GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 25, 2004, 01:54:26 ; Search time 2.78648 Seconds
(without alignments)
912.596 Million cell updates/sec

Title: US-09-593-793A-113_COPY_367_375
Perfect score: 48
Sequence: 1 GATCLSHSV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	48	100.0	123	4	AAM01228	Aam01228 P553S spl
2	48	100.0	123	4	AAU69873	Aau69873 Human pro
3	48	100.0	123	4	ABU71764	Abu71764 Prostate
4	48	100.0	123	5	ABB95333	Abb95333 Human P55
5	48	100.0	123	6	ABR54445	Abr54445 Prostate
6	48	100.0	123	7	ADB14156	Adb14156 Human pro
7	48	100.0	231	3	AAY54369	Aay54369 Amino aci
8	48	100.0	246	6	ABR54582	Abr54582 Prostate
9	48	100.0	246	7	ADB14478	Adb14478 Human P50
10	48	100.0	252	4	AAE01423	Aae01423 Human sec
11	48	100.0	255	2	AAW85068	Aaw85068 Protein e
12	48	100.0	255	3	AAB29268	Aab29268 Human pro
13	48	100.0	255	4	AAU04205	Aau04205 Prostate-
14	48	100.0	255	5	AAO19084	Aao19084 Human pro
15	48	100.0	305	6	ABR54583	Abr54583 Prostate
16	48	100.0	305	7	ADB14479	Adb14479 Human P50
17	48	100.0	355	6	ABR54578	Abr54578 Prostate
18	48	100.0	355	7	ADB14461	Adb14461 Fusion co
19	48	100.0	553	2	AAW69385	Aaw69385 Prostate
20	48	100.0	553	2	AAW71869	Aaw71869 Amino aci
21	48	100.0	553	3	AAY82002	Aay82002 Human imm
22	48	100.0	553	3	AAB28527	Aab28527 Protein e
23	48	100.0	553	3	ABG94411	Abg94411 Human pro
24	48	100.0	553	4	AAM01117	Aam01117 Human pro
25	48	100.0	553	4	AAU69763	Aau69763 Human pro
26	48	100.0	553	4	AAB74800	Aab74800 Prostate
27	48	100.0	553	4	AAG99002	Aag99002 Human pro
28	48	100.0	553	4	AAG62150	Aag62150 Human P50
29	48	100.0	553	4	ABU71653	Abu71653 Prostate
30	48	100.0	553	4	AAU04961	Aau04961 Human pro
31	48	100.0	553	5	AAU10324	Aau10324 Human PRO
32	48	100.0	553	5	ABB95222	Abb95222 Human L1-
33	48	100.0	553	5	AAU82643	Aau82643 Human bre
34	48	100.0	553	5	ABG61900	Abg61900 Prostate
35	48	100.0	553	5	ABG76665	Abg76665 Prostate
36	48	100.0	553	5	ABB77575	Abb77575 Human mas
37	48	100.0	553	6	ABR54334	Abr54334 Prostate
38	48	100.0	553	7	ADB13563	Adb13563 Human pro
39	48	100.0	595	4	AAM01318	Aam01318 Alpha pre
40	48	100.0	710	5	AAM50661	Aam50661 Thioredox
41	48	100.0	1079	4	AAB74830	Aab74830 Prostate
42	48	100.0	1079	4	ABU71860	Abu71860 Prostate
43	44	91.7	20	4	AAM01145	Aam01145 Human pro
44	44	91.7	20	4	AAU69790	Aau69790 Human pro
45	44	91.7	20	4	AAG99030	Aag99030 Human pro

ALIGNMENTS

RESULT 1
AAM01228
ID AAM01228 standard; protein; 123 AA.
XX
AC AAM01228;
XX
DT 04-OCT-2001 (first entry)
XX
DE P553S splice variant amino acid P553S-6.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US001574.
XX
PR 14-JAN-2000; 2000US-00483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for diagnosing,
PT monitoring and treating prostate cancer in a patient and for use in
PT vaccines.
XX
PS Claim 2; Page 463-464; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated T cells
CC prepared using (I) or (II) are used treat cancer in a patient. (I) and
CC the antibodies are also used in the detection of cancer in a patient. The
CC cancer that is diagnosed or treated is particularly prostate cancer. (I)
CC and (II) can be used in vaccines. The antibodies or (I) can be used for
CC monitoring the progression of cancer in a patient. (I) and (II) can also
CC be used to improve diagnostic and therapeutic methods for prostate
CC cancer. They can indicate the level of metastasis as well as the prostate

CC volume. AAH93357 to AAH93944 and AAM01115 to AAM01318 represent
CC polynucleotide and amino acid sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 123 AA;

Query Match 100.0%; Score 48; DB 4; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 2
AAU69873
ID AAU69873 standard; protein; 123 AA.
XX
AC AAU69873;
XX
DT 30-JAN-2002 (first entry)
XX
DE Human prostate cDNA encoded protein #67.
XX
KW Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX
OS Homo sapiens.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US009919.
XX
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2001-639232/73.
DR N-PSDB; AAS64041.
XX
PT New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer.
XX
PS Claim 2; Page 486-487; 579pp; English.
XX
CC The invention relates to isolated prostate-specific polynucleotides,
CC polypeptides, fusion proteins of the polypeptides, antibodies raised
CC against the polypeptides (or antigenic epitopes derived from them) and
CC antigen-presenting cells expressing the polypeptides. The antibodies are
CC useful for detecting the presence of cancer, especially prostate cancer.
CC The polypeptides, polynucleotides and the antigen-presenting cells are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein, and for inhibiting the development of cancer especially prostate
CC cancer. Compositions comprising the polynucleotide and/or polypeptide are
CC useful for stimulating an immune response, and for treating cancer. The
CC oligonucleotide is useful for detecting cancer. The present sequence is a
CC prostate specific polypeptide of the invention
XX
SQ Sequence 123 AA;

Query Match 100.0%; Score 48; DB 4; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 3
ABU71764
ID ABU71764 standard; protein; 123 AA.
XX
AC ABU71764;
XX
DT 10-JUN-2003 (first entry)
XX
DE Prostate cancer associated protein #33.
XX
KW Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
KW immunogen; cancer; prostate specific antigen; PSA;
KW prostatic acid phosphatase; PAP; prostate specific membrane antigen;
KW PSMA.
XX
OS Homo sapiens.
XX
PN US2002192763-A1.
XX
PD 19-DEC-2002.
XX
PF 29-JUN-2001; 2001US-00895793.
XX
PR 04-OCT-1999; 99US-0157455P.
PR 04-OCT-2000; 2000US-00679272.
PR 28-MAR-2001; 2001US-00822827.
XX
PA (XUJJ/) XU J.
PA (DILL/) DILLON D C.
PA (MITC/) MITCHAM J L.
PA (HARL/) HARLOCKER S L.
PA (JIAN/) JIANG Y.
PA (KALO/) KALOS M D.
PA (FANG/) FANGER G R.
PA (RETT/) RETTER M W.
PA (STOL/) STOLK J A.
PA (DAYC/) DAY C H.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (LISX/) LI S X.
PA (WANG/) WANG A.
PA (SKEI/) SKEIKY Y A W.
PA (HEPL/) HEPLER W T.
PA (HEND/) HENDERSON R A.
PA (HURA/) HURAL J.
PA (MCNE/) MCNEILL P D.
PA (HOUG/) HOUGHTON R L.
PA (DBAS/) Y DE BASSOLS C V.
PA (FOYT/) FOY T M.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA, Hural J;
PI Mcneill PD, Houghton RL, Y De BassolsCV, Foy TM;
XX
DR WPI; 2001-245062/25.
XX
PT Prostate specific protein and its encoding polynucleotide, useful for the
PT treatment and diagnosis of prostate cancer.
XX
PS Example 13; SEQ ID NO 706; 85pp; English.

XX
CC The invention describes a fusion protein comprising at least one amino
CC acid sequence of immunogenic portions of any of the 3 sequences not
CC defined in the specification, or sequences having at least 70 or 90 %
CC sequence identity to any one of the 35 sequences defined in the USPTO web
CC site, which is encoded by any of the 4 nucleotide sequences not defined
CC in the specification. The fusion protein, composition and methods are
CC useful for diagnosing, preventing and/or treating cancer, particularly
CC prostate cancer. The proteins are useful as markers to indicate the
CC presence or absence of cancer. This is the amino acid sequence of a
CC prostate cancer therapy associated protein. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20020192763
XX
SQ Sequence 123 AA;

Query Match 100.0%; Score 48; DB 4; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 4
ABB95333
ID ABB95333 standard; protein; 123 AA.
XX
AC ABB95333;
XX
DT 19-JUL-2002 (first entry)
XX
DE Human P553S splice variant encoded protein SEQ ID NO 706.
XX
KW Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN US2002022248-A1.
XX
PD 21-FEB-2002.
XX
PF 12-JAN-2001; 2001US-00759143.
XX
PR 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 10-FEB-1998; 98US-00020956.
PR 25-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 15-JAN-1999; 99US-00232149.
PR 09-APR-1999; 99US-00288946.
PR 13-JUL-1999; 99US-00352616.
PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 10-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
XX
PA (XUJJ/) XU J.
PA (DILL/) DILLON D C.
PA (MITC/) MITCHAM J L.
PA (HARL/) HARLOCKER S L.
PA (JIAN/) JIANG Y.
PA (KALO/) KALOS M D.
PA (FANG/) FANGER G R.
PA (RETT/) RETTER M W.
PA (STOL/) STOLK J A.
PA (DAYC/) DAY C H.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (LISX/) LI S X.
PA (WANG/) WANG A.
PA (SKEI/) SKEIKY Y A W.
PA (HEPL/) HEPLER W T.
PA (HEND/) HENDERSON R A.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2002-255649/30.
XX
PT New prostate-specific polynucleotides for diagnosing and treating
PT diseases, in particular prostate cancer, and as markers for the
PT progression of cancer.
XX
PS Claim 2; SEQ ID NO 706; 87pp; English.
XX
CC The present invention provides prostate-specific coding sequences and
CC their encoded proteins. These can be used in the diagnosis and treatment
CC of cancers, particularly prostate cancer. The present sequence is a
CC protein described in the invention
XX
SQ Sequence 123 AA;

Query Match 100.0%; Score 48; DB 5; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 5
ABR54445
ID ABR54445 standard; protein; 123 AA.
XX
AC ABR54445;
XX
DT 28-AUG-2003 (first entry)
XX
DE Prostate tumour specific protein sequence SEQ ID 706.
XX
KW Cytostatic; gene therapy; prostate-specific protein; PSP; human;
KW immune response; prostate cancer.
XX
OS Homo sapiens.
XX
PN WO200289747-A2.
XX
PD 14-NOV-2002.
XX
PF 09-MAY-2002; 2002WO-US014753.
XX
PR 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Henderson RA;
PI Kalos MD, Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS;
PI Carter D, Li SX, Wang A, Skeiky YAW, Hepler WT, Hural J;

PI Mcneill PD, Houghton RL, Vinals Y De BassolsC, Foy TM, Watanabe Y;
PI Deng T;
XX
DR WPI; 2003-167130/16.
XX
PT New prostate-specific proteins and genes, useful in gene therapy,
PT particularly for stimulating an immune response in a patient, or treating
PT prostate cancer in a patient, as well as for diagnosing prostate cancer
PT in a patient.
XX
PS Example 13; Page 555; 691pp; English.
XX
CC The present invention relates to novel prostate-specific proteins (PSP)
CC and their coding sequences. The PSPs and their coding sequences are
CC useful for stimulating an immune response in a patient, or for treating
CC prostate cancer in a patient and for determining, detecting or diagnosing
CC the presence of a cancer in a patient. The present sequence was used to
CC illustrate the invention
XX
SQ Sequence 123 AA;

Query Match 100.0%; Score 48; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 6
ADB14156
ID ADB14156 standard; protein; 123 AA.
XX
AC ADB14156;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human prostate specific cDNA P553S splice variant #4 protein.
XX
KW Human; prostate specific cDNA; cytostatic; immunostimulant; gene therapy;
KW cell therapy; vaccine; T-cell epitope;
KW class I major histocompatibility complex allele; MHC; prostate cancer;
KW tumour; antigen presenting cell.
XX
OS Homo sapiens.
XX
PN US2003185830-A1.
XX
PD 02-OCT-2003.
XX
PF 12-NOV-2002; 2002US-00294025.
XX
PR 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 09-FEB-1998; 98US-00020956.
PR 25-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 15-JAN-1999; 99US-00232149.
PR 09-APR-1999; 99US-00288946.
PR 13-JUL-1999; 99US-00352616.
PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
PR 12-JAN-2001; 2001US-00759143.
PR 09-FEB-2001; 2001US-00780669.
PR 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
PR 09-MAY-2002; 2002US-00144678.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Stolk JA, Kalos MD;
XX
DR WPI; 2003-756193/71.
DR N-PSDB; ADB14155.
XX
PT New isolated polypeptide for use in a vaccine for stimulating an immune
PT response, or for treating or diagnosis cancer, preferably prostate
PT cancer.
XX
PS Example 13; Page; 101pp; English.
XX
CC The invention relates to an isolated polypeptide comprising no more than
CC 11-542 amino acids of ADB13563 comprising a sequence ADB14487. The
CC peptides comprise a fragment ADB13563 of that contain naturally processed
CC T-cell epitopes for 3 class I major histocompatibility complex (MHC)
CC alleles. ADB13563 is a polypeptide encoded by a human prostate specific
CC cDNA, one of 648 disclosed as new. Also included are nucleic acids
CC encoding the proteins and peptides, expression vectors, a host cell
CC transformed with the vector, an isolated antibody (or antigen binding
CC fragment) that specifically binds to the protein or peptide, detecting
CC the presence of a cancer in a patient (comprising contacting a patient
CC sample with a binding agent that binds to the peptides or a polypeptide
CC appearing as ADB13558, detecting the amount of polypeptide that binds to
CC the agent and comparing the amount of polypeptide to a predetermined cut-
CC off value to determine the presence of cancer), a fusion protein
CC comprising the peptides or proteins, stimulating or expanding T cells
CC specific for a tumour protein comprising contacting T cells with the
CC peptides or the isolated T cell population, treating prostate cancer in a
CC patient comprising administering a composition comprising the peptides,
CC nucleic acids, antibodies or compounds, determining the presence of a
CC cancer in a patient and treating prostate cancer in a patient comprising
CC incubating cluster of differentiation (CD4+) and/or CD8+ T cells isolated
CC from a patient with the peptides or antigen presenting cells that express
CC (the peptides so that the T cells proliferate, and administering the
CC proliferated T cells to the patient. The peptides (or an oligonucleotide
CC that hybridises to nucleic acid encoding them), is used to detect the
CC presence of cancer in a patient. The peptides, nucleic acids encoding, or
CC antigen-presenting cells expressing the nucleic acid, are used to
CC stimulate or expand T cells specific for a tumour protein. The peptides,
CC nucleic acids, antibodies, fusion proteins, T cell populations or antigen
CC presenting cells are used to stimulate an immune response or treat
CC prostate cancer in a patient. The present sequence is a prostate specific
CC protein of the invention. Note: Except where otherwise indicated, the
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=20030185830.
XX
SQ Sequence 123 AA;

Query Match 100.0%; Score 48; DB 7; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 7
AAY54369
ID AAY54369 standard; protein; 231 AA.

```
XX
AC   AAY54369;
XX
DT   06-APR-2000  (first entry)
XX
DE   Amino acid sequence of a prostate cancer-associated protein.
XX
KW   Prostate cancer-associated gene; Incyte clone 1864683; bone cancer;
KW   cell proliferation; cancer; adrenal gland cancer; bladder cancer;
KW   prostate cancer.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .47
FT                   /note= "signal peptide"
FT   Modified-site   100
FT                   /note= "potential casein kinase II phosphorylation site"
FT   Modified-site   142
FT                   /note= "potential casein kinase II phosphorylation site"
FT   Modified-site   147
FT                   /note= "potential protein kinase C phosphorylation site"
FT   Domain          188. .209
FT                   /note= "potential transmembrane domain"
XX
PN   WO9967384-A2.
XX
PD   29-DEC-1999.
XX
PF   15-JUN-1999;   99WO-US013524.
XX
PR   22-JUN-1998;   98US-00102615.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Walker MG,  Volkmuth W,  Klingler TM,  Sprinzak EA;
XX
DR   WPI; 2000-126631/11.
DR   N-PSDB; AAZ45677.
XX
PT   Identifying biomolecules for the diagnosis and treatment of diseases
PT   associated with cell-proliferation.
XX
PS   Claim 10; Page 50-51; 52pp; English.
XX
CC   The present sequence is encoded by a prostate cancer-associated gene,
CC   whose cDNA is represented by Incyte clone number 1864683. The cDNA
CC   sequence is used in the method of the invention. The specification
CC   describes a method for identifying biomolecules for the diagnosis or
CC   treatment of diseases associated with cell proliferation. The method
CC   comprises examining polynucleotides, consisting of prostate cancer-
CC   specific genes, and genes of unknown function, expressed in cDNA
CC   libraries. The patterns of both gene sets are compared to identify genes
CC   of unknown function with similar expression patterns to the prostate
CC   cancer-specific genes. The biomolecules identified by the method form
CC   pharmaceutical compositions useful for the diagnosis and treatment of
CC   diseases associated with cell proliferation. Such diseases include cancer
CC   of the adrenal gland, bladder and bone, but especially prostate cancer.
CC   The method may also be applied using other disease- specific genes. The
CC   prostate cancer-specific genes facilitate the diagnosis and treatment of
CC   cell proliferation disorders
XX
SQ   Sequence 231 AA;

  Query Match             100.0%;  Score 48;  DB 3;  Length 231;
  Best Local Similarity   100.0%;  Pred. No. 1.1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              |||||||||
Db         45 GATCLSHSV 53
```

```
RESULT 8
ABR54582
ID   ABR54582 standard; protein; 246 AA.
XX
AC   ABR54582;
XX
DT   28-AUG-2003  (first entry)
XX
DE   Prostate tumour-related protein SEQ ID 1028.
XX
KW   Cytostatic; gene therapy; prostate-specific protein; PSP; human;
KW   immune response; prostate cancer.
XX
OS   Homo sapiens.
XX
PN   WO200289747-A2.
XX
PD   14-NOV-2002.
XX
PF   09-MAY-2002; 2002WO-US014753.
XX
PR   09-MAY-2001; 2001US-00852911.
PR   29-JUN-2001; 2001US-00895814.
PR   10-DEC-2001; 2001US-00012896.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Henderson RA;
PI   Kalos MD,  Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS;
PI   Carter D,  Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Hural J;
PI   Mcneill PD,  Houghton RL,  Vinals Y De BassolsC,  Foy TM,  Watanabe Y;
PI   Deng T;
XX
DR   WPI; 2003-167130/16.
XX
PT   New prostate-specific proteins and genes, useful in gene therapy,
PT   particularly for stimulating an immune response in a patient, or treating
PT   prostate cancer in a patient, as well as for diagnosing prostate cancer
PT   in a patient.
XX
PS   Claim 2; Page 688-689; 691pp; English.
XX
CC   The present invention relates to novel prostate-specific proteins (PSP)
CC   and their coding sequences. The PSPs and their coding sequences are
CC   useful for stimulating an immune response in a patient, or for treating
CC   prostate cancer in a patient and for determining, detecting or diagnosing
CC   the presence of a cancer in a patient. The present sequence was used to
CC   illustrate the invention
XX
SQ   Sequence 246 AA;

  Query Match             100.0%;  Score 48;  DB 6;  Length 246;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              |||||||||
Db         60 GATCLSHSV 68
```

```
RESULT 9
ADB14478
ID   ADB14478 standard; protein; 246 AA.
XX
AC   ADB14478;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human P501S deletion mutant, P501S D.
XX
KW   Human; prostate specific cDNA; cytostatic; immunostimulant; gene therapy;
KW   cell therapy; vaccine; T-cell epitope;
KW   class I major histocompatibility complex allele; MHC; prostate cancer;
```

KW tumour; antigen presenting cell.
XX
OS Homo sapiens.
XX
PN US2003185830-A1.
XX
PD 02-OCT-2003.
XX
PF 12-NOV-2002; 2002US-00294025.
XX
PR 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 09-FEB-1998; 98US-00020956.
PR 25-FEB-1998; 98US-00030607.
PR 14-JUL-1998; 98US-00115453.
PR 23-SEP-1998; 98US-00159812.
PR 15-JAN-1999; 99US-00232149.
PR 09-APR-1999; 99US-00288946.
PR 13-JUL-1999; 99US-00352616.
PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
PR 14-JAN-2000; 2000US-00483672.
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
PR 12-JAN-2001; 2001US-00759143.
PR 09-FEB-2001; 2001US-00780669.
PR 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
PR 09-MAY-2002; 2002US-00144678.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Stolk JA, Kalos MD;
XX
DR WPI; 2003-756193/71.
DR N-PSDB; ADB14476.
XX
PT New isolated polypeptide for use in a vaccine for stimulating an immune
PT response, or for treating or diagnosis cancer, preferably prostate
PT cancer.
XX
PS Example 49; Page; 101pp; English.
XX
CC The invention relates to an isolated polypeptide comprising no more than
CC 11-542 amino acids of ADB13563 comprising a sequence ADB14487. The
CC peptides comprise a fragment ADB13563 of that contain naturally processed
CC T-cell epitopes for 3 class I major histocompatibility complex (MHC)
CC alleles. ADB13563 is a polypeptide encoded by a human prostate specific
CC cDNA, one of 648 disclosed as new. Also included are nucleic acids
CC encoding the proteins and peptides, expression vectors, a host cell
CC transformed with the vector, an isolated antibody (or antigen binding
CC fragment) that specifically binds to the protein or peptide, detecting
CC the presence of a cancer in a patient (comprising contacting a patient
CC sample with a binding agent that binds to the peptides or a polypeptide
CC appearing as ADB13558, detecting the amount of polypeptide that binds to
CC the agent and comparing the amount of polypeptide to a predetermined cut-
CC off value to determine the presence of cancer), a fusion protein
CC comprising the peptides or proteins, stimulating or expanding T cells
CC specific for a tumour protein comprising contacting T cells with the
CC peptides or the isolated T cell population, treating prostate cancer in a
CC patient comprising administering a composition comprising the peptides,
CC nucleic acids, antibodies or compounds, determining the presence of a
CC cancer in a patient and treating prostate cancer in a patient comprising
CC incubating cluster of differentiation (CD4+) and/or CD8+ T cells isolated
CC from a patient with the peptides or antigen presenting cells that express
CC (the peptides so that the T cells proliferate, and administering the
CC proliferated T cells to the patient. The peptides (or an oligonucleotide
CC that hybridises to nucleic acid encoding them), is used to detect the
CC presence of cancer in a patient. The peptides, nucleic acids encoding, or
CC antigen-presenting cells expressing the nucleic acid, are used to
CC stimulate or expand T cells specific for a tumour protein. The peptides,
CC nucleic acids, antibodies, fusion proteins, T cell populations or antigen
CC presenting cells are used to stimulate an immune response or treat
CC prostate cancer in a patient. The present sequence is a prostate specific
CC protein of the invention. Note: Except where otherwise indicated, the
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=20030185830.
XX
SQ Sequence 246 AA;

Query Match 100.0%; Score 48; DB 7; Length 246;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 60 GATCLSHSV 68

RESULT 10
AAE01423
ID AAE01423 standard; protein; 252 AA.
XX
AC AAE01423;
XX
DT 18-JUL-2001 (first entry)
XX
DE Human secreted protein variant, SEQ ID NO:147.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnerary; binding partner identification;
KW gene therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 162
FT /note= "Corresponds to any of the naturally occuring L-
FT amino acids"
FT Misc-difference 222
FT /note= "Corresponds to any of the naturally occuring L-
FT amino acids"
FT Misc-difference 236
FT /note= "Corresponds to any of the naturally occuring L-
FT amino acids"
FT Misc-difference 240
FT /note= "Corresponds to any of the naturally occuring L-
FT amino acids"
XX
PN WO200134629-A1.
XX
PD 17-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US030654.
XX
PR 12-NOV-1999; 99US-0164835P.
PR 27-JUL-2000; 2000US-0221142P.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Wei P, Fiscella M, Baker KP;
XX
DR WPI; 2001-308779/32.
XX
PT New nucleic acid encoding one of 21 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions, such
PT as autoimmune disease and cancer, and used as a food additive or
PT preservative.
XX
PS Disclosure; Page 476-477; 490pp; English.
XX
CC AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted
CC protein genes, and AAE01352-AAE01413 represent the proteins they encode.
CC AAE01415-AAE01433 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 21 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
CC present sequence represents a human secreted protein variant referred to
CC in the disclosure of the invention
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 48; DB 4; Length 252;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 11
AAW85068
ID AAW85068 standard; protein; 255 AA.
XX
AC AAW85068;
XX
DT 12-FEB-1999 (first entry)
XX
DE Protein encoded by the consensus sequence of the PS108 gene.
XX
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW drug screening; gene therapy.
XX
OS Homo sapiens.
XX
PN WO9850567-A1.
XX
PD 12-NOV-1998.
XX
PF 01-MAY-1998; 98WO-US008930.
XX
PR 02-MAY-1997; 97US-00850713.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;
PI Granados EN, Hodges SC, Klass MR, Kratochvil JD, Roberts-Rapp L;
PI Russell JC, Stroupe SD;
XX
DR WPI; 1999-034731/03.
DR N-PSDB; AAV71181.
XX
PT New isolated prostate-specific polynucleotides - used to develop products
PT for the diagnosis and treatment of prostate diseases, e.g. benign
PT hyperplasia, prostatic or prostate cancer.
XX
PS Claim 17; Page 99-100; 122pp; English.
XX
CC The present sequence is encoded by the consensus sequence for a PS108
CC gene. The sequence is derived from overlapping clones AAV71166-79. The
CC clone sequences are PS108 gene-specific. They are used in the method of
CC the invention. The specification describes a method for detecting the
CC presence of a target PS108 polynucleotide in a test sample. The method
CC comprises contacting the test sample with at least 1 PS108-specific
CC polynucleotide or complement, and detecting the presence of the target
CC PS108 polynucleotide. The products can be used for detecting, diagnosing,
CC staging, monitoring, prognosticating, in vivo imaging, preventing or
CC treating, or determining predisposition to diseases or conditions of the
CC prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC products can be used in drug screening and gene therapy
XX
SQ Sequence 255 AA;

Query Match 100.0%; Score 48; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 69 GATCLSHSV 77

RESULT 12
AAB29268
ID AAB29268 standard; protein; 255 AA.
XX
AC AAB29268;
XX
DT 07-FEB-2001 (first entry)
XX
DE Human prostate-related PS108 partial protein sequence SEQ ID NO: 36.
XX
KW Human; prostate cancer; PS108; antibody; tumour; metastasis.
XX
OS Homo sapiens.
XX
PN US6130043-A.
XX
PD 10-OCT-2000.
XX
PF 01-MAY-1998; 98US-00071710.
XX
PR 02-MAY-1997; 97US-00850713.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Cohen M, Colpitts TL, Friedman PN, Gordon J, Granados EN;
PI Billing-Medel PA, Klass MR, Roberts-Rapp L, Stroupe SD, Yu H;
PI Kratochvil JD, Russell JC, Hodges SC;
XX

DR WPI; 2000-655655/63.
XX
PT Methods for detecting target prostate-specific polynucleotides or
PT diseases of the prostate (e.g. prostate cancer), comprising detecting the
PT presence of any of PS108 nucleic acid sequences in a test sample.
XX
PS Example 1; Col 85-88; 55pp; English.
XX
CC The present invention is related to a number of partial coding and
CC protein sequences for the human prostate tissue protein PS108. These
CC sequences can be used in the diagnosis and prognosis of prostate
CC diseases, particularly prostate cancer. They can also be used to produce
CC antibodies which can be used in treatment. The present sequence is one of
CC the PS108 partial protein sequences
XX
SQ Sequence 255 AA;

Query Match 100.0%; Score 48; DB 3; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
||||||||| |
Db 69 GATCLSHSV 77

RESULT 13
AAU04205
ID AAU04205 standard; protein; 255 AA.
XX
AC AAU04205;
XX
DT 23-OCT-2001 (first entry)
XX
DE Prostate-specific PS108 polypeptide #1.
XX
KW Prostate; PS108; immunogen; drug screening; image localisation;
KW diagnostic; therapeutic; prostate tissue disease; cancer; metastasis;
KW expressed sequence tag; EST.
XX
OS Homo sapiens.
XX
PN US6252047-B1.
XX
PD 26-JUN-2001.
XX
PF 15-MAR-2000; 2000US-00525397.
XX
PR 02-MAY-1997; 97US-00850713.
PR 01-MAY-1998; 98US-00071710.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Billingel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;
PI Granados EN, Hodges SC, Klass MR, Kratochvil JD, Roberts-Rapp L;
PI Russell JC, Stroupe SD, Yu H;
XX
DR WPI; 2001-424488/45.
DR N-PSDB; AAS07155.
XX
PT Novel PS108 polypeptide useful in assays for detecting antibodies to
PT prostate tissue, and as immunogens to produce PS108 antibodies.
XX
PS Claim 1; Col 85-87; 55pp; English.
XX
CC The sequence represents the amino acid sequence of prostate-specific
CC PS108 #1, encoded by PS108 expressed sequence tag (EST) consensus
CC sequence. The sequence was produced from overlapping PS108 ESTs sequences
CC to produce a full length consensus sequence. This sequence was then used
CC to produce the PS108 polypeptide which is useful in assays for detecting
CC antibodies to prostate tissue, and as an immunogen to produce antibodies.
CC The polypeptide is useful for screening compounds which specifically bind
CC to the polypeptide and for screening for drugs, compounds, or any other
CC agent which can be used to treat diseases associated with PS108. The
CC antibody is useful to detect, or for image localisation of, PS108 antigen
CC in a patient, for detecting or diagnosing a disease or condition, as
CC delivery agents for therapeutic agents as well as for diagnostic tests
CC and for screening for diseases or conditions associated with PS108,
CC especially cancer. The antibody is also useful for generating chimeric
CC antibodies for therapeutic use, for inhibiting the biological activity of
CC PS108, in therapy (for e.g. to treat prostate tissue disease including
CC prostate cancer and its metastases), and to detect the presence of any
CC polypeptide in a test sample which shares one or more antigenic
CC determinants with the PS108 polypeptide
XX
SQ Sequence 255 AA;

Query Match 100.0%; Score 48; DB 4; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
||||||||| |
Db 69 GATCLSHSV 77

RESULT 14
AAO19084
ID AAO19084 standard; protein; 255 AA.
XX
AC AAO19084;
XX
DT 22-NOV-2002 (first entry)
XX
DE Human prostate-specific PS108 protein SEQ ID NO: 36.
XX
KW Human; PS108; prostate cancer; prostate specific sequence; prostate;
KW cytostatic; gene therapy.
XX
OS Homo sapiens.
XX
PN US2002086301-A1.
XX
PD 04-JUL-2002.
XX
PF 25-APR-2001; 2001US-00841894.
XX
PR 02-MAY-1997; 97US-00850713.
PR 01-MAY-1998; 98US-00071710.
XX
PA (BILL/) BILLINGEL P A.
PA (COHE/) COHEN M.
PA (COLP/) COLPITTS T L.
PA (FRIE/) FRIEDMAN P N.
PA (GORD/) GORDON J.
PA (GRAN/) GRANADOS E N.
PA (HODG/) HODGES S C.
PA (KLAS/) KLASS M R.
PA (KRAT/) KRATOCHVIL J D.
PA (ROBE/) ROBERTS-RAPP L A.
PA (RUSS/) RUSSELL J C.
PA (STRO/) STROUPE S D.
PA (YUHH/) YU H.
XX
PI Billingel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;
PI Granados EN, Hodges SC, Klass MR, Kratochvil JD, Roberts-Rapp LA;
PI Russell JC, Stroupe SD, Yu H;
XX
DR WPI; 2002-665428/71.
XX
PT Novel PS108 polynucleotides derived from the PS108 gene, useful for
PT detecting, diagnosing, staging, monitoring, prognosticating, in vivo
PT imaging, preventing or treating diseases and conditions of prostate e.g.,
PT prostate cancer.
XX
PS Claim 17; Page 46; 59pp; English.

XX
CC The present invention relates to polynucleotide sequences derived from
CC the human PS108 gene and capable of selectively hybridising to the PS108
CC gene. The sequences can be used to detect the presence of a target PS108
CC polynucleotide in a test sample, and to detect, monitor, prevent or treat
CC diseases and conditions of the prostate, such as prostate cancer. The
CC present sequence is a PS108 protein fragment
XX
SQ Sequence 255 AA;

Query Match 100.0%; Score 48; DB 5; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
||||||||| |
Db 69 GATCLSHSV 77

RESULT 15
ABR54583
ID ABR54583 standard; protein; 305 AA.
XX
AC ABR54583;
XX
DT 28-AUG-2003 (first entry)
XX
DE Prostate tumour-related protein SEQ ID 1029.
XX
KW Cytostatic; gene therapy; prostate-specific protein; PSP; human;
KW immune response; prostate cancer.
XX
OS Homo sapiens.
XX
PN WO200289747-A2.
XX
PD 14-NOV-2002.
XX
PF 09-MAY-2002; 2002WO-US014753.
XX
PR 09-MAY-2001; 2001US-00852911.
PR 29-JUN-2001; 2001US-00895814.
PR 10-DEC-2001; 2001US-00012896.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Henderson RA;
PI Kalos MD, Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS;
PI Carter D, Li SX, Wang A, Skeiky YAW, Hepler WT, Hural J;
PI Mcneill PD, Houghton RL, Vinals Y De BassolsC, Foy TM, Watanabe Y;
PI Deng T;
XX
DR WPI; 2003-167130/16.
XX
PT New prostate-specific proteins and genes, useful in gene therapy,
PT particularly for stimulating an immune response in a patient, or treating
PT prostate cancer in a patient, as well as for diagnosing prostate cancer
PT in a patient.
XX
PS Claim 2; Page 689-690; 691pp; English.
XX
CC The present invention relates to novel prostate-specific proteins (PSP)
CC and their coding sequences. The PSPs and their coding sequences are
CC useful for stimulating an immune response in a patient, or for treating
CC prostate cancer in a patient and for determining, detecting or diagnosing
CC the presence of a cancer in a patient. The present sequence was used to
CC illustrate the invention
XX
SQ Sequence 305 AA;

Query Match 100.0%; Score 48; DB 6; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
||||||||| |
Db 119 GATCLSHSV 127

Search completed: August 25, 2004, 02:27:08
Job time : 2.78648 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 25, 2004, 02:24:18 ; Search time 2.64235 Seconds
(without alignments)
1070.372 Million cell updates/sec

Title: US-09-593-793A-113_COPY_367_375
Perfect score: 48
Sequence: 1 GATCLSHSV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	48	100.0	22	14	US-10-010-940-569	Sequence 569, App
2	48	100.0	123	9	US-09-759-143-706	Sequence 706, App
3	48	100.0	123	9	US-09-780-669-706	Sequence 706, App
4	48	100.0	123	9	US-09-822-827-706	Sequence 706, App
5	48	100.0	123	9	US-09-895-793-706	Sequence 706, App
6	48	100.0	123	9	US-09-895-814-706	Sequence 706, App
7	48	100.0	123	13	US-10-012-896-706	Sequence 706, App
8	48	100.0	123	14	US-10-144-678A-706	Sequence 706, App
9	48	100.0	123	14	US-10-294-025-706	Sequence 706, App
10	48	100.0	246	14	US-10-144-678A-1028	Sequence 1028, Ap
11	48	100.0	246	14	US-10-294-025-1028	Sequence 1028, Ap
12	48	100.0	255	12	US-09-841-894-36	Sequence 36, Appl
13	48	100.0	305	14	US-10-144-678A-1029	Sequence 1029, Ap
14	48	100.0	305	14	US-10-294-025-1029	Sequence 1029, Ap
15	48	100.0	355	13	US-10-012-896-1011	Sequence 1011, Ap
16	48	100.0	355	14	US-10-144-678A-1011	Sequence 1011, Ap
17	48	100.0	355	14	US-10-294-025-1011	Sequence 1011, Ap
18	48	100.0	553	9	US-09-745-288-101	Sequence 101, App
19	48	100.0	553	9	US-09-838-785-2	Sequence 2, Appli
20	48	100.0	553	9	US-09-759-143-113	Sequence 113, App
21	48	100.0	553	9	US-09-780-669-113	Sequence 113, App
22	48	100.0	553	9	US-09-030-606-113	Sequence 113, App
23	48	100.0	553	9	US-09-822-827-113	Sequence 113, App
24	48	100.0	553	9	US-09-115-453-113	Sequence 113, App
25	48	100.0	553	9	US-09-232-880-113	Sequence 113, App
26	48	100.0	553	9	US-09-895-793-113	Sequence 113, App
27	48	100.0	553	9	US-09-895-814-113	Sequence 113, App
28	48	100.0	553	12	US-10-453-919-101	Sequence 101, App
29	48	100.0	553	13	US-10-012-896-113	Sequence 113, App
30	48	100.0	553	14	US-10-010-940-113	Sequence 113, App
31	48	100.0	553	14	US-10-144-678A-113	Sequence 113, App
32	48	100.0	553	14	US-10-005-907-13	Sequence 13, Appl
33	48	100.0	553	14	US-10-294-025-113	Sequence 113, App
34	48	100.0	553	15	US-10-295-027-548	Sequence 548, App
35	48	100.0	553	15	US-10-295-027-902	Sequence 902, App
36	48	100.0	553	16	US-10-688-838-113	Sequence 113, App
37	48	100.0	710	14	US-10-296-770-4	Sequence 4, Appli
38	48	100.0	1079	9	US-09-822-827-947	Sequence 947, App
39	48	100.0	1079	9	US-09-895-793-947	Sequence 947, App
40	44	91.7	20	9	US-09-759-143-490	Sequence 490, App
41	44	91.7	20	9	US-09-780-669-490	Sequence 490, App
42	44	91.7	20	9	US-09-822-827-490	Sequence 490, App
43	44	91.7	20	9	US-09-895-793-490	Sequence 490, App
44	44	91.7	20	9	US-09-895-814-490	Sequence 490, App
45	44	91.7	20	13	US-10-012-896-490	Sequence 490, App

ALIGNMENTS

RESULT 1
US-10-010-940-569
; Sequence 569, Application US/10010940
; Publication No. US20030088062A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427D3
; CURRENT APPLICATION NUMBER: US/10/010,940
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 569
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-010-940-569

Query Match 100.0%; Score 48; DB 14; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 9 GATCLSHSV 17

RESULT 2
US-09-759-143-706
; Sequence 706, Application US/09759143
; Patent No. US20020022248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 706
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-143-706

Query Match 100.0%; Score 48; DB 9; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 3
US-09-780-669-706
; Sequence 706, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780,669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 706
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-780-669-706

Query Match 100.0%; Score 48; DB 9; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 4
US-09-822-827-706
; Sequence 706, Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 706
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-822-827-706

Query Match 100.0%; Score 48; DB 9; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 5
US-09-895-793-706
; Sequence 706, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 706
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-706

Query Match 100.0%; Score 48; DB 9; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 6
US-09-895-814-706
; Sequence 706, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 706
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-814-706

Query Match 100.0%; Score 48; DB 9; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 7
US-10-012-896-706
; Sequence 706, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 706
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-706

Query Match 100.0%; Score 48; DB 13; Length 123;
Best Local Similarity 100.0%; Pred. No. 0.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 45 GATCLSHSV 53

RESULT 8
US-10-144-678A-706
; Sequence 706, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Hepler, William T.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.

```
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  Vinals y de Bassols, Carlota
;  APPLICANT:  Foy, Teresa M.
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Deng, Ta
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C28
;  CURRENT APPLICATION NUMBER: US/10/144,678A
;  CURRENT FILING DATE:  2002-08-12
;  NUMBER OF SEQ ID NOS: 1033
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 706
;   LENGTH: 123
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-144-678A-706

  Query Match             100.0%;  Score 48;  DB 14;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 0.6;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              |||||||||
Db         45 GATCLSHSV 53


RESULT 9
US-10-294-025-706
; Sequence 706, Application US/10294025
; Publication No. US20030185830A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Kalos, Michael D.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C29
;  CURRENT APPLICATION NUMBER: US/10/294,025
;  CURRENT FILING DATE:  2002-11-12
;  NUMBER OF SEQ ID NOS: 1038
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 706
;   LENGTH: 123
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-294-025-706

  Query Match             100.0%;  Score 48;  DB 14;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 0.6;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              |||||||||
Db         45 GATCLSHSV 53


RESULT 10
US-10-144-678A-1028
; Sequence 1028, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A. W.
;  APPLICANT:  Hepler, William T.
;  APPLICANT:  Hural, John
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  Vinals y de Bassols, Carlota
;  APPLICANT:  Foy, Teresa M.
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Deng, Ta
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C28
;  CURRENT APPLICATION NUMBER: US/10/144,678A
;  CURRENT FILING DATE:  2002-08-12
;  NUMBER OF SEQ ID NOS: 1033
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1028
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-144-678A-1028

  Query Match             100.0%;  Score 48;  DB 14;  Length 246;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              |||||||||
Db         60 GATCLSHSV 68


RESULT 11
US-10-294-025-1028
; Sequence 1028, Application US/10294025
; Publication No. US20030185830A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Kalos, Michael D.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C29
;  CURRENT APPLICATION NUMBER: US/10/294,025
;  CURRENT FILING DATE:  2002-11-12
;  NUMBER OF SEQ ID NOS: 1038
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1028
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-294-025-1028

  Query Match             100.0%;  Score 48;  DB 14;  Length 246;
  Best Local Similarity   100.0%;  Pred. No. 1.2;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              |||||||||
Db         60 GATCLSHSV 68


RESULT 12
US-09-841-894-36
; Sequence 36, Application US/09841894
; Publication No. US20020086301A1
;   GENERAL INFORMATION:
;        APPLICANT: BILLING-MEDEL, PATRICIA
;                   COHEN, MAURICE
;                   COLPITTS, TRACEY L.
```

; FRIEDMAN, PAULA N.
; GORDON, JULIAN
; GRANADOS, EDWARD N.
; HODGES, STEVEN C.
; KLASS, MICHAEL R.
; KRATOCHVIL, JON D.
; ROBERTS-RAPP, LISA
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; FOR DETECTING DISEASES OF THE PROSTATE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,894
; FILING DATE: 25-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,710
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6083.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 255 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: None
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-841-894-36

Query Match 100.0%; Score 48; DB 12; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 69 GATCLSHSV 77

RESULT 13
US-10-144-678A-1029
; Sequence 1029, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Hepler, William T.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals y de Bassols, Carlota
; APPLICANT: Foy, Teresa M.
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Deng, Ta
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C28
; CURRENT APPLICATION NUMBER: US/10/144,678A
; CURRENT FILING DATE: 2002-08-12
; NUMBER OF SEQ ID NOS: 1033
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1029
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-678A-1029

Query Match 100.0%; Score 48; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 119 GATCLSHSV 127

RESULT 14
US-10-294-025-1029
; Sequence 1029, Application US/10294025
; Publication No. US20030185830A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C29
; CURRENT APPLICATION NUMBER: US/10/294,025
; CURRENT FILING DATE: 2002-11-12
; NUMBER OF SEQ ID NOS: 1038
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1029
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-294-025-1029

Query Match 100.0%; Score 48; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
|||||||||
Db 119 GATCLSHSV 127

RESULT 15
US-10-012-896-1011
; Sequence 1011, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.

```
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Hepler, William T.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Hural, John
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  Vinals de Bassols, Carlota
;  APPLICANT:  Foy, Teresa
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Wantanabe, Yoshihiro
;  APPLICANT:  Meagher, Madeleine Joy
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C27
;  CURRENT APPLICATION NUMBER: US/10/012,896
;  CURRENT FILING DATE:  2001-12-10
;  NUMBER OF SEQ ID NOS: 1011
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1011
;   LENGTH: 355
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-012-896-1011

  Query Match             100.0%;  Score 48;  DB 13;  Length 355;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              |||||||||
Db        185 GATCLSHSV 193


Search completed: August 25, 2004, 02:34:07
Job time : 2.64235 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 25, 2004, 02:00:53 ; Search time 0.800712 Seconds
(without alignments)
1081.193 Million cell updates/sec

Title: US-09-593-793A-113_COPY_367_375
Perfect score: 48
Sequence: 1 GATCLSHSV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	77.1	310	2	A55768	asparaginyl-peptid
2	37	77.1	669	2	T32512	hypothetical prote
3	36	75.0	308	2	JC7125	epidermal growth f
4	36	75.0	1208	2	T34469	hypothetical prote
5	35	72.9	511	1	S31308	aldehyde dehydroge
6	34	70.8	156	2	S54619	hypothetical prote
7	34	70.8	311	2	F69820	conserved hypothet
8	34	70.8	898	2	A40114	fasciclin II precu
9	34	70.8	1545	2	T26589	hypothetical prote
10	33	68.8	170	2	AF2621	polypeptide deform
11	33	68.8	170	2	F97403	polypeptide deform
12	33	68.8	187	2	AF3542	formylmethionine d
13	33	68.8	302	2	S59853	DNA-binding protei
14	33	68.8	357	2	T41716	conserved hypothet
15	33	68.8	383	2	S53716	delta-like homeoti
16	33	68.8	385	2	A54785	preadipocyte facto
17	33	68.8	385	2	S53718	homeotic protein d
18	33	68.8	443	1	QQBYPT	hypothetical prote
19	33	68.8	627	2	A83408	phenazine biosynth
20	33	68.8	1001	2	S30385	G9a protein - huma
21	33	68.8	1076	2	D82083	carbamoyl-phosphat
22	33	68.8	1216	2	T26104	hypothetical prote
23	33	68.8	1227	2	T49963	hypothetical prote
24	32	66.7	64	2	E69172	sensory transducti
25	32	66.7	148	2	D81852	conserved hypothet
26	32	66.7	174	2	AB1888	hypothetical prote
27	32	66.7	246	2	JQ1472	trypsin (EC 3.4.21
28	32	66.7	246	2	JQ1471	trypsin (EC 3.4.21
29	32	66.7	262	1	S17705	indole-3-glycerol-
30	32	66.7	270	2	T46856	indole-3-glycerol-
31	32	66.7	277	1	D35115	indole-3-glycerol-
32	32	66.7	294	2	A72362	cell division prot
33	32	66.7	468	2	T49682	hypothetical prote
34	32	66.7	638	2	AC1123	B. subtilis IolD p
35	32	66.7	638	2	AE1483	B. subtilis IolD p
36	32	66.7	705	2	A35621	spore germination
37	32	66.7	775	2	T37837	probable signal tr
38	32	66.7	1106	2	T18739	hypothetical prote
39	32	66.7	1672	2	T46237	hypothetical prote
40	32	66.7	2023	2	T13154	polycomb protein e
41	32	66.7	2731	1	VFIHJH	genome polyprotein
42	32	66.7	2733	2	S15760	genome polyprotein
43	31	64.6	99	2	F82824	hypothetical prote
44	31	64.6	163	2	H83168	conserved hypothet
45	31	64.6	166	2	AB0371	conserved hypothet

ALIGNMENTS

RESULT 1
A55768
asparaginyl-peptide amidohydrolase (EC 3.5.1.-) - pig
N;Alternate names: asparagine amidohydrolase
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 05-Nov-1999
C;Accession: A55768
R;Stewart, A.E.; Arfin, S.M.; Bradshaw, R.A.
J. Biol. Chem. 270, 25-28, 1995
A;Title: The sequence of porcine protein NH-2-terminal asparagine amidohydrolase. A new
A;Reference number: A55768; MUID:95113832; PMID:7814382
A;Accession: A55768
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-310 <STE>
A;Cross-references: GB:U17062; NID:g595949; PIDN:AAA65019.1; PID:g595950
C;Keywords: hydrolase
F;2-310/Product: asparaginyl-peptide amidohydrolase #status predicted <MAT>

Query Match 77.1%; Score 37; DB 2; Length 310;
Best Local Similarity 85.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GATCLSH 7
              |||||:|
Db         86 GATCLTH 92
```

RESULT 2
T32512
hypothetical protein C44B12.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32512
R;Tin-Wollam, A.
submitted to the EMBL Data Library, December 1997
A;Description: The sequence of C. elegans cosmid C44B12.
A;Reference number: Z21183
A;Accession: T32512
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-669 <TIN>
A;Cross-references: EMBL:AF036692; PIDN:AAB88327.1; GSPDB:GN00022; CESP:C44B12.4
A;Experimental source: strain Bristol N2; clone C44B12
C;Genetics:
A;Gene: CESP:C44B12.4
A;Map position: 4
A;Introns: 35/3; 89/2; 133/3; 164/1; 213/3; 288/2; 327/2; 361/3; 380/1; 448/2; 638/2

Query Match 77.1%; Score 37; DB 2; Length 669;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 GATCLSHSV 9
              |  |||||:
Db         62 GTLCLSHSI 70
```

RESULT 3
JC7125
epidermal growth factor-like homeotic protein PREF-1 - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 02-Aug-2002
C;Accession: JC7125
R;Fahrenkrug, S.C.; Freking, B.A.; Smith, T.P.L.
Biochem. Biophys. Res. Commun. 264, 662-667, 1999
A;Title: Genomic organization and genetic mapping of the bovine PREF-1 gene.
A;Reference number: JC7125; MUID:20012729; PMID:10543989
A;Accession: JC7125
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-308 <FAH>
A;Cross-references: GB:AF181831; NID:g5932411
C;Genetics:
A;Gene: Dlk
A;Map position: 21 (77.5 cM)
C;Superfamily: preadipocyte factor 1; EGF homology

```
  Query Match             75.0%;  Score 36;  DB 2;  Length 308;
  Best Local Similarity   75.0%;  Pred. No. 21;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GATCLSHS 8
              | ||| ||
Db        220 GGTCLQHS 227
```

RESULT 4
T34469
hypothetical protein W03A3.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34469
R;Miller, N.
submitted to the EMBL Data Library, February 1996
A;Description: The sequence of C. elegans cosmid W03A3.
A;Reference number: Z21531
A;Accession: T34469
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1208 <MIL>
A;Cross-references: EMBL:U50184; PIDN:AAB93325.1; GSPDB:GN00021; CESP:W03A3.2
A;Experimental source: strain Bristol N2; clone W03A3
C;Genetics:
A;Gene: CESP:W03A3.2
A;Map position: 3
A;Introns: 93/3; 119/3; 197/3; 221/1; 267/2; 302/2; 555/1; 589/3; 666/1; 692/1; 744/3; 9

```
  Query Match             75.0%;  Score 36;  DB 2;  Length 1208;
  Best Local Similarity   85.7%;  Pred. No. 69;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GATCLSH 7
              | |||||
Db        275 GVTCLSH 281
```

RESULT 5
S31308
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 precursor, mitochondrial - yeast (Saccharomy
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002
C;Accession: S31308
R;Thielen, J.
submitted to the EMBL Data Library, October 1992
A;Reference number: S31308
A;Accession: S31308
A;Molecule type: DNA
A;Residues: 1-511 <THI>
A;Cross-references: EMBL:Z17314; NID:g3361; PIDN:CAA78962.1; PID:g3362
C;Comment: This form is probably mitochondrial.
C;Genetics:
A;Gene: ALD2
C;Function:
A;Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
A;Note: enzymes with this activity are involved in diverse metabolic pathways in various
C;Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C;Keywords: alcohol metabolism; mitochondrion; NAD; oxidoreductase
F;1-21/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F;22-511/Product: aldehyde dehydrogenase (NAD+) 2 #status predicted <MAT>
F;80-351/Domain: aldehyde dehydrogenase homology <ALDD>
F;297,331/Active site: Glu, Cys #status predicted
F;483/Binding site: NAD (Cys) #status predicted

```
  Query Match             72.9%;  Score 35;  DB 1;  Length 511;
  Best Local Similarity   85.7%;  Pred. No. 51;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GATCLSH 7
              || ||||
Db        258 GAACLSH 264
```

RESULT 6
S54619
hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C;Accession: S54619; S66879
R;de Haan, M.; Maarse, A.C.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54617
A;Accession: S54619
A;Molecule type: DNA
A;Residues: 1-156 <DEH>
A;Cross-references: EMBL:X87331; NID:g1041652; PIDN:CAA60762.1; PID:g829123
R;de Haan, M.; Grivell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66877
A;Accession: S66879
A;Molecule type: DNA
A;Residues: 1-156 <DEW>
A;Cross-references: EMBL:Z74920; NID:g1420109; PIDN:CAA99201.1; PID:g1420111; MIPS:YOR01
A;Experimental source: strain S288C
C;Genetics:
A;Cross-references: SGD:S0005539
A;Map position: 15R
C;Superfamily: hypothetical protein YOR013w

```
  Query Match             70.8%;  Score 34;  DB 2;  Length 156;
  Best Local Similarity   85.7%;  Pred. No. 29;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GATCLSH 7
              ||||| |
Db         23 GATCLLH 29
```

RESULT 7
F69820
conserved hypothetical protein yhbB - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: F69820
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

```
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F69820
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-311 <KUN>
A;Cross-references: GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CAB12720.1; PID:g2633215
A;Experimental source: strain 168
C;Genetics:
A;Gene: yhbB

  Query Match             70.8%;  Score 34;  DB 2;  Length 311;
  Best Local Similarity   55.6%;  Pred. No. 52;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              | ||||:::
Db         72 GTTCLSYTI 80


RESULT 8
A40114
fasciclin II precursor - American bird grasshopper
C;Species: Schistocerca americana (American bird grasshopper)
C;Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 17-Mar-2000
C;Accession: A40114; B31817
R;Harrelson, A.L.; Goodman, C.S.
Science 242, 700-708, 1988
A;Title: Growth cone guidance in insects: fasciclin II is a member of the immunoglobulin
A;Reference number: A40114; MUID:89043938; PMID:3187519
A;Accession: A40114
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 1-898 <HAR>
R;Snow, P.M.; Zinn, K.; Harrelson, A.L.; McAllister, L.; Schilling, J.; Bastiani, M.J.;
Proc. Natl. Acad. Sci. U.S.A. 85, 5291-5295, 1988
A;Title: Characterization and cloning of fasciclin I and fasciclin II glycoproteins in t
A;Reference number: A94202; MUID:88276943; PMID:2839842
A;Accession: B31817
A;Molecule type: mRNA
A;Residues: 423-436 <SNO>
A;Cross-references: GB:J03789; NID:g160848; PID:g160849
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C;Keywords: transmembrane protein

  Query Match             70.8%;  Score 34;  DB 2;  Length 898;
  Best Local Similarity   71.4%;  Pred. No. 1.3e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCLSHSV 9
              |||:||:
Db        449 TCLAHSI 455


RESULT 9
T26589
hypothetical protein Y32B12B.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26589
R;Gardner, A.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z20239
A;Accession: T26589
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1545 <WIL>
A;Cross-references: EMBL:AL031632; PIDN:CAA21009.1; GSPDB:GN00023; CESP:Y32B12B.4
A;Experimental source: clone Y32B12B
C;Genetics:
A;Gene: CESP:Y32B12B.4
A;Map position: 5
A;Introns: 49/2; 311/2; 879/2; 1116/2; 1257/3; 1404/2; 1545/2

  Query Match             70.8%;  Score 34;  DB 2;  Length 1545;
  Best Local Similarity   62.5%;  Pred. No. 2e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 ATCLSHSV 9
              ||| :||:
Db       1494 ATCFTHSI 1501


RESULT 10
AF2621
polypeptide deformylase def [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AF2621
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AF2621
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-170 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL41388.1; PID:g17738706; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: def
A;Map position: circular chromosome
C;Superfamily: polypeptide deformylase

  Query Match             68.8%;  Score 33;  DB 2;  Length 170;
  Best Local Similarity   62.5%;  Pred. No. 49;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 ATCLSHSV 9
              |||| | :
Db        131 ATCLQHEI 138


RESULT 11
F97403
polypeptide deformylase (AF213822) [imported] - Agrobacterium tumefaciens (strain C58, (
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: F97403
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.,
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tur
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: F97403
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-170 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK86183.1; PID:g15155278; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_640
```

```
A;Map position: circular chromosome
C;Superfamily: polypeptide deformylase

  Query Match             68.8%;  Score 33;  DB 2;  Length 170;
  Best Local Similarity   62.5%;  Pred. No. 49;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 ATCLSHSV 9
              |||| | :
Db        131 ATCLQHEI 138


RESULT 12
AF3542
formylmethionine deformylase (EC 3.5.1.31) [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 19-Apr-2002
C;Accession: AF3542
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AF3542
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-187 <KUR>
A;Cross-references: GB:AE008918; PIDN:AAL53505.1; PID:g17984410; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEII0264
A;Map position: II
C;Superfamily: polypeptide deformylase
C;Keywords: hydrolase

  Query Match             68.8%;  Score 33;  DB 2;  Length 187;
  Best Local Similarity   62.5%;  Pred. No. 53;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 ATCLSHSV 9
              |||| | :
Db        143 ATCLQHEI 150


RESULT 13
S59853
DNA-binding protein Dof3 - maize (fragment)
C;Species: Zea mays (maize)
C;Date: 15-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 29-Oct-1999
C;Accession: S59853; S59851
R;Shuichi, Y.
submitted to the EMBL Data Library, June 1994
A;Reference number: S59852
A;Accession: S59853
A;Molecule type: mRNA
A;Residues: 1-302 <SHU>
A;Cross-references: EMBL:X79935; NID:g1061307; PIDN:CAA56288.1; PID:g1061308
R;Yanagisawa, S.
Nucleic Acids Res. 23, 3403-3410, 1995
A;Title: A novel DNA-binding domain that may form a single zinc finger motif.
A;Reference number: S59850; MUID:96032831; PMID:7567449
A;Accession: S59851
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 27-99 <YAN>
A;Cross-references: EMBL:X79935
C;Keywords: zinc finger

  Query Match             68.8%;  Score 33;  DB 2;  Length 302;
  Best Local Similarity   66.7%;  Pred. No. 79;
  Matches    6;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
                || |  |||
Db        176 GAACYPHSV 184


RESULT 14
T41716
conserved hypothetical protein SPAC821.05 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T41716
R;Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z22012
A;Accession: T41716
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-357 <RIE>
A;Cross-references: EMBL:AL121770; PIDN:CAB57439.1; GSPDB:GN00066; SPDB:SPAC821.05
A;Experimental source: strain 972h-; cosmid c821
C;Genetics:
A;Gene: SPDB:SPAC821.05
A;Map position: 1
A;Introns: 9/2; 45/1

  Query Match             68.8%;  Score 33;  DB 2;  Length 357;
  Best Local Similarity   75.0%;  Pred. No. 91;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 ATCLSHSV 9
              |||| ||:
Db        200 ATCLLHSL 207


RESULT 15
S53716
delta-like homeotic protein dlk, long splice form precursor - human
N;Alternate names: fetal antigen 1 (FA1)
N;Contains: delta-like homeotic protein dlk, short splice form
C;Species: Homo sapiens (man)
C;Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 02-Aug-2002
C;Accession: S53716; S53717; S71068; B45484; S48713; A44549; S31973; S31974
R;Lee, Y.L.; Helman, L.; Hoffman, T.; Laborda, J.
Biochim. Biophys. Acta 1261, 223-232, 1995
A;Title: dlk, pG2 and Pref-1 mRNAs encode similar proteins belonging to the EGF-like sup
A;Reference number: S53716; MUID:95226449; PMID:7711066
A;Accession: S53716
A;Molecule type: mRNA
A;Residues: 1-383 <LEE1>
A;Cross-references: EMBL:U15979; NID:g562105; PIDN:AAA75364.1; PID:g562106
A;Note: the authors translated the codon CAG for residue 46 as His and CCT for residue 4
A;Note: the sequence in GenBank entry HSU15979 has a 3 base deletion mutation in the reg
A;Accession: S53717
A;Molecule type: mRNA
A;Residues: 1-228,302-383 <LEE2>
A;Cross-references: EMBL:U15981; NID:g562109; PIDN:AAA75365.1; PID:g562110
R;Laborda, J.
submitted to the EMBL Data Library, October 1994
A;Reference number: S71068
A;Accession: S71068
A;Molecule type: mRNA
A;Residues: 1-344,346-383 <LAB1>
A;Cross-references: EMBL:U15979; NID:g562105; PIDN:AAA75364.1; PID:g562106
R;Laborda, J.; Sausville, E.A.; Hoffman, T.; Notario, V.
J. Biol. Chem. 268, 3817-3820, 1993
A;Title: dlk, a putative mammalian homeotic gene differentially expressed in small cell
A;Reference number: A45484; MUID:93179372; PMID:8095043
A;Accession: B45484
A;Molecule type: mRNA
A;Residues: 1-45,'HV',48-383 <LAB2>
A;Cross-references: EMBL:Z12172; NID:g38478; PIDN:CAA78163.1; PID:g38479
A;Note: sequence extracted from NCBI backbone (NCBIP:125735)
R;Jensen, C.H.; Krogh, T.N.; Hojrup, P.; Clausen, P.P.; Skjodt, K.; Larsson, L.I.; Enghi
Eur. J. Biochem. 225, 83-92, 1994
```

```
A;Title: Protein structure of fetal antigen 1 (FA1). A novel circulating human epidermal
 pG2.
A;Reference number: S48713; MUID:95010145; PMID:7925474
A;Accession: S48713
A;Molecule type: protein
A;Residues: 24-107,'D',109-282 <JEN>
R;Hojrup, P.; Jensen, C.H.; Skjodt, K.; Teisner, B.
Protein Sci. 2(Suppl.1), 259S, 1993
A;Title: Primary structure of human fetal antigen 1 (FA1), a putative homeotic glycoprot
A;Reference number: A44549
A;Accession: A44549
A;Status: preliminary
A;Molecule type: protein
A;Residues: 24-107,'VASM',112-283 <HOJ>
C;Genetics:
A;Gene: GDB:DLK1; dlk; FA1; PG2; PREF-1
A;Cross-references: GDB:9958854; OMIM:176290
A;Map position: 14q32-14q32
C;Superfamily: preadipocyte factor 1; EGF homology
C;Keywords: alternative splicing; glycoprotein; transmembrane protein
F;1-23/Domain: signal sequence #status predicted <SIG>
F;21-54/Domain: EGF homology #status atypical <EGF1>
F;24-383/Product: delta-like homeotic protein dlk, long splice form #status predicted <M
F;24-282/Product: fetal antigen 1 #status experimental <MATA>
F;24-228,302-383/Product: delta-like homeotic protein dlk, short splice form #status pre
F;57-85/Domain: EGF homology <EGF2>
F;92-124/Domain: EGF homology <EGF3>
F;131-167/Domain: EGF homology <EGF4>
F;174-205/Domain: EGF homology <EGF5>
F;212-244/Domain: EGF homology <EGF6>
F;305-327/Domain: transmembrane #status predicted <TRM>
F;94,214/Binding site: carbohydrate (Ser) (covalent) #status experimental
F;100/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;134/Binding site: carbohydrate (Asn) (covalent) #status absent
F;143/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;163,251,260/Binding site: carbohydrate (Ser) (covalent) (partial) #status experimental
F;165,172/Binding site: carbohydrate (Asn) (covalent) (partial) #status atypical
F;222/Binding site: carbohydrate (Thr) (covalent) (partial) #status experimental

  Query Match             68.8%;  Score 33;  DB 2;  Length 383;
  Best Local Similarity   62.5%;  Pred. No. 97;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GATCLSHS 8
              | ||| |:
Db        220 GGTCLQHT 227


Search completed: August 25, 2004, 02:28:40
Job time : 2.80071 secs
```

OM protein - protein search, using sw model

Run on: August 25, 2004, 01:58:37 ; Search time 0.480427 Seconds
(without alignments)
975.448 Million cell updates/sec

Title: US-09-593-793A-113_COPY_367_375
Perfect score: 48
Sequence: 1 GATCLSHSV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	81.2	1376	1	CRBH_HUMAN	P82279 homo sapien
2	37	77.1	309	1	PNAD_HUMAN	Q96ab6 homo sapien
3	37	77.1	309	1	PNAD_MOUSE	Q64311 mus musculu
4	37	77.1	309	1	PNAD_PIG	Q28955 sus scrofa
5	35	72.9	511	1	DHAY_YEAST	P32872 saccharomyc
6	34	70.8	110	1	TNG2_HUMAN	P56847 homo sapien
7	34	70.8	898	1	FAS2_SCHAM	P22648 schistocerc
8	34	70.8	2209	1	KNTC_HUMAN	P50748 homo sapien
9	33	68.8	170	1	DEF_AGRT5	Q8uid1 agrobacteri
10	33	68.8	174	1	DEF_RHIME	Q92sh6 rhizobium m
11	33	68.8	175	1	DEF_BRUME	Q8ydb4 brucella me
12	33	68.8	176	1	DEF_RHILO	Q98d52 rhizobium l
13	33	68.8	383	1	DLK_HUMAN	P80370 homo sapien
14	33	68.8	385	1	DLK_MOUSE	Q09163 mus musculu
15	33	68.8	443	1	YEY2_YEAST	P10356 saccharomyc
16	33	68.8	637	1	PHZB_PSECL	Q51519 pseudomonas
17	33	68.8	637	1	PHZE_PSEFL	Q51791 pseudomonas
18	33	68.8	1076	1	CARB_VIBCH	Q9kph9 vibrio chol
19	33	68.8	1077	1	CARB_VIBPA	Q87sf3 vibrio para
20	33	68.8	1077	1	CARB_VIBVU	Q8dem2 vibrio vuln
21	33	68.8	1210	1	BAT8_HUMAN	Q96kq7 homo sapien
22	33	68.8	1263	1	BAT8_MOUSE	Q9z148 mus musculu
23	32	66.7	148	1	YF93_NEIMA	Q51134 neisseria m
24	32	66.7	246	1	TRYA_RAT	P32821 rattus norv
25	32	66.7	246	1	TRYB_RAT	P32822 rattus norv
26	32	66.7	262	1	TRPC_AZOBR	P26938 azospirillu
27	32	66.7	265	1	TRPC_XANAC	Q8pq47 xanthomonas
28	32	66.7	270	1	TRPC_RHOSH	Q9zfa7 rhodobacter
29	32	66.7	277	1	TRPC_PSEPK	Q88qr6 pseudomonas
30	32	66.7	277	1	TRPC_PSEPU	P20578 pseudomonas
31	32	66.7	278	1	TRPC_PSESM	Q88a03 pseudomonas
32	32	66.7	337	1	CT86_HUMAN	Q9bz19 homo sapien
33	32	66.7	705	1	GUN6_DICDI	P22699 dictyosteli
34	32	66.7	735	1	SALF_HUMAN	Q9y6q2 homo sapien
35	32	66.7	7059	1	R1AB_CVBQ	Q8v6w7 b replicase
36	32	66.7	7094	1	R1AB_CVBEN	Q91a29 b replicase
37	32	66.7	7094	1	R1AB_CVBLU	Q8v439 b replicase
38	32	66.7	7094	1	R1AB_CVBM	Q66198 b replicase
39	32	66.7	7124	1	R1AB_CVM2	Q9pya3 m replicase
40	32	66.7	7176	1	R1AB_CVMA5	P16342 m replicase
41	32	66.7	7180	1	R1AB_CVMJH	P19751 m replicase
42	31	64.6	174	1	DEF_BRAJA	Q89bn9 bradyrhizob
43	31	64.6	177	1	YCX8_CHLVU	O20162 chlorella v
44	31	64.6	246	1	TRY1_RAT	P00762 rattus norv
45	31	64.6	248	1	TRY3_CHICK	Q90629 gallus gall

ALIGNMENTS

```
RESULT 1
CRBH_HUMAN
ID   CRBH_HUMAN     STANDARD;      PRT;  1376 AA.
AC   P82279;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Crumbs protein homolog 1 precursor.
GN   CRB1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANTS RP12 VAL-161; TRP-250; MET-745;
RP   CYS-764; TYR-948; THR-1041 AND PRO-1071.
RC   TISSUE=Fetal brain, and Retina;
RX   MEDLINE=99438399; PubMed=10508521;
RA   den Hollander A.I., ten Brink J.B., de Kok Y.J.M., van Soest S.,
RA   van den Born L.I., van Driel M.A., van de Pol D.J.R., Payne A.M.,
RA   Bhattacharya S.S., Kellner U., Hoyng C.B., Westerveld A.,
RA   Brunner H.G., Bleeker-Wagemakers E.M., Deutman A.F.,
RA   Heckenlively J.R., Cremers F.P.M., Bergen A.A.B.;
RT   "Mutations in a human homologue of Drosophila crumbs cause retinitis
RT   pigmentosa (RP12).";
RL   Nat. Genet. 23:217-221(1999).
RN   [2]
RP   VARIANT LCA ARG-1100, AND VARIANT RP ARG-1181.
RX   MEDLINE=21303018; PubMed=11389483;
RA   den Hollander A.I., Heckenlively J.R., van den Born L.I.,
RA   de Kok Y.J.M., van der Velde-Visser S.D., Kellner U., Jurklies B.,
RA   van Schooneveld M.J., Blankenagel A., Rohrschneider K., Wissinger B.,
RA   Cruysberg J.R.M., Deutman A.F., Brunner H.G., Apfelstedt-Sylla E.,
RA   Hoyng C.B., Cremers F.P.M.;
RT   "Leber congenital amaurosis and retinitis pigmentosa with coats-like
RT   exudative vasculopathy are associated with mutations in the crumbs
RT   homologue 1 (CRB1) gene.";
RL   Am. J. Hum. Genet. 69:198-203(2001).
CC   -!- FUNCTION: May be involved in cell-cell interaction in neuronal
CC       development of the retina.
CC   -!- SUBCELLULAR LOCATION: Extracellular (Potential).
CC   -!- TISSUE SPECIFICITY: Preferential expression in retina, also
CC       expressed in brain and fetal brain.
CC   -!- DISEASE: Defects in CRB1 are the cause of retinitis pigmentosa
CC       type 12 (RP12) [MIM:600105]. RP12 is an autosomal recessive
CC       condition which is characterized by night blindness from early
CC       childhood and progressive visual field loss. There is general loss
CC       of retinal pigment epithelium throughout the retina and patients
CC       experience severe visual impairment before the age of twenty.
CC   -!- DISEASE: Defects in CRB1 are a cause of Leber congenital amaurosis
CC       (LCA) [MIM:204000]. LCA designates a group of autosomal recessive
CC       retinal dystrophies that represent the most common genetic causes
CC       of congenital visual impairment in infants and children.
CC   -!- SIMILARITY: Contains 15 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 laminin G-like domains.
CC   -!- DATABASE: NAME=Mutations of the CRB1 gene;
```

```
CC       NOTE=Retina International's Scientific Newsletter;
CC       WWW="http://www.retina-international.com/sci-news/crb1mut.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF154671; AAF01361.1; -.
DR   HSSP; P08709; 1BF9.
DR   Genew; HGNC:2343; CRB1.
DR   MIM; 604210; -.
DR   MIM; 204000; -.
DR   MIM; 600105; -.
DR   GO; GO:0007267; P:cell-cell signaling; TAS.
DR   GO; GO:0007163; P:establishment and/or maintenance of cell po. . .; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00008; EGF; 16.
DR   Pfam; PF00054; laminin_G; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00179; EGF_CA; 8.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 10.
DR   PROSITE; PS00022; EGF_1; 15.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS50026; EGF_3; 19.
DR   PROSITE; PS01187; EGF_CA; 7.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   EGF-like domain; Glycoprotein; Repeat; Signal; Disease mutation;
KW   Retinitis pigmentosa; Vision.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26   1376       CRUMBS PROTEIN HOMOLOG 1.
FT   DOMAIN      70    108       EGF-LIKE 1.
FT   DOMAIN     110    146       EGF-LIKE 2.
FT   DOMAIN     148    184       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     186    222       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     224    260       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     304    337       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     339    395       EGF-LIKE 7.
FT   DOMAIN     441    481       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     485    670       LAMININ G-LIKE 1.
FT   DOMAIN     672    708       EGF-LIKE 9.
FT   DOMAIN     714    885       LAMININ G-LIKE 2.
FT   DOMAIN     887    923       EGF-LIKE 10.
FT   DOMAIN     950   1137       LAMININ G-LIKE 3.
FT   DOMAIN    1139   1175       EGF-LIKE 11.
FT   DOMAIN    1177   1212       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1214   1250       EGF-LIKE 13.
FT   DOMAIN    1255   1295       EGF-LIKE 14.
FT   DOMAIN    1297   1333       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DISULFID    74     85       POTENTIAL.
FT   DISULFID    79     96       POTENTIAL.
FT   DISULFID    98    107       POTENTIAL.
FT   DISULFID   114    125       POTENTIAL.
FT   DISULFID   119    134       POTENTIAL.
FT   DISULFID   136    145       POTENTIAL.
FT   DISULFID   152    163       POTENTIAL.
FT   DISULFID   157    172       POTENTIAL.
FT   DISULFID   174    183       POTENTIAL.
FT   DISULFID   190    201       POTENTIAL.
FT   DISULFID   195    210       POTENTIAL.
FT   DISULFID   212    221       POTENTIAL.
FT   DISULFID   228    239       POTENTIAL.
FT   DISULFID   233    248       POTENTIAL.
FT   DISULFID   250    259       POTENTIAL.
FT   DISULFID   305    316       POTENTIAL.
FT   DISULFID   310    325       POTENTIAL.
FT   DISULFID   327    336       POTENTIAL.
FT   DISULFID   343    354       POTENTIAL.
FT   DISULFID   348    383       POTENTIAL.
FT   DISULFID   385    394       POTENTIAL.
FT   DISULFID   445    456       POTENTIAL.
FT   DISULFID   450    469       POTENTIAL.
FT   DISULFID   471    480       POTENTIAL.
FT   DISULFID   676    687       POTENTIAL.
FT   DISULFID   681    696       POTENTIAL.
FT   DISULFID   698    707       POTENTIAL.
FT   DISULFID   891    902       POTENTIAL.
FT   DISULFID   896    911       POTENTIAL.
FT   DISULFID   913    922       POTENTIAL.
FT   DISULFID  1143   1154       POTENTIAL.
FT   DISULFID  1148   1163       POTENTIAL.
FT   DISULFID  1165   1174       POTENTIAL.
FT   DISULFID  1181   1191       POTENTIAL.
FT   DISULFID  1186   1200       POTENTIAL.
FT   DISULFID  1202   1211       POTENTIAL.
FT   DISULFID  1218   1229       POTENTIAL.
FT   DISULFID  1223   1238       POTENTIAL.
FT   DISULFID  1240   1249       POTENTIAL.
FT   DISULFID  1259   1274       POTENTIAL.
FT   DISULFID  1268   1283       POTENTIAL.
FT   DISULFID  1285   1294       POTENTIAL.
FT   DISULFID  1301   1312       POTENTIAL.
FT   DISULFID  1306   1321       POTENTIAL.
FT   DISULFID  1323   1332       POTENTIAL.
FT   CARBOHYD    30     30       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    41     41       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    42     42       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   215    215       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   287    287       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   313    313       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   322    322       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   418    418       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   427    427       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   453    453       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   550    550       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   561    561       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   657    657       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   757    757       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   871    871       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   880    880       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   968    968       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   975    975       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1000   1000       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1190   1190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1243   1243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1265   1265       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1273   1273       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT    161    161       A -> V (in RP12).
FT                               /FTId=VAR_011641.
FT   VARIANT    250    250       C -> W (in RP12).
FT                               /FTId=VAR_011642.
FT   VARIANT    745    745       T -> M (in RP12).
FT                               /FTId=VAR_011643.
FT   VARIANT    764    764       R -> C (in RP12).
FT                               /FTId=VAR_011644.
FT   VARIANT    948    948       C -> Y (in RP12).
FT                               /FTId=VAR_011645.
FT   VARIANT   1041   1041       M -> T (in RP12).
FT                               /FTId=VAR_011646.
FT   VARIANT   1071   1071       L -> P (in RP12).
FT                               /FTId=VAR_011647.
FT   VARIANT   1100   1100       I -> R (in LCA).
FT                               /FTId=VAR_011648.
FT   VARIANT   1181   1181       C -> R (in RP12; with coat-like exudative
FT                               vasculopathy).
FT                               /FTId=VAR_011649.
```

```
  Query Match          81.2%;  Score 39;  DB 1;  Length 1376;
  Best Local Similarity  75.0%;  Pred. No. 7.8;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHS 8
              ||||:||:
Db       1226 GATCISHT 1233


RESULT 2
PNAD_HUMAN
ID   PNAD_HUMAN     STANDARD;      PRT;   309 AA.
AC   Q96AB6;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein N-terminal asparagine amidohydrolase (EC 3.5.1.-) (Protein
DE   NH2-terminal asparagine deamidase) (N-terminal Asn amidase) (NTN-
DE   amidase) (PNAD) (Protein NH2-terminal asparagine amidohydrolase)
DE   (PNAA).
GN   NTAN1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Side chain deamidation of N-terminal asparagine residues
CC       to aspartate. Required for the ubiquitin-dependent turnover of
CC       intracellular proteins that initiate with Met-Asn. These proteins
CC       are acetylated on the retained initiator methionine and can
CC       subsequently be modified by the removal of N-acetyl methionine by
CC       acylaminoacid hydrolase (AAH). Conversion of the resulting N-
CC       terminal asparagine to aspartate by PNAD renders the protein
CC       susceptible to arginylation, polyubiquitinylation and degradation
CC       as specified by the N-end rule. This enzyme does not act on
CC       substrates with internal or C-terminal asparagines and does not
CC       act on glutamine residues in any position (By similarity).
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; BC017336; AAH17336.1; -.
KW   Hydrolase.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   309 AA;  34546 MW;  75FE6EA52347E485 CRC64;

  Query Match          77.1%;  Score 37;  DB 1;  Length 309;
  Best Local Similarity  85.7%;  Pred. No. 3.9;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSH 7
              |||||:|
Db         85 GATCLTH 91


RESULT 3
PNAD_MOUSE
ID   PNAD_MOUSE     STANDARD;      PRT;   309 AA.
AC   Q64311;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein N-terminal asparagine amidohydrolase (EC 3.5.1.-) (Protein
DE   NH2-terminal asparagine deamidase) (N-terminal Asn amidase) (NTN-
DE   amidase) (PNAD) (Protein NH2-terminal asparagine amidohydrolase)
DE   (PNAA).
GN   NTAN1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C129;
RX   MEDLINE=97067079; PubMed=8910481;
RA   Grigoryev S., Stewart A.E., Kwon Y.T., Arfin S.M., Bradshaw R.A.,
RA   Jenkins N.A., Copeland N.G., Varshavsky A.;
RT   "A mouse amidase specific for N-terminal asparagine. The gene, the
RT   enzyme, and their function in the N-end rule pathway.";
RL   J. Biol. Chem. 271:28521-28532(1996).
CC   -!- FUNCTION: Side chain deamidation of N-terminal asparagine residues
CC       to aspartate. Required for the ubiquitin-dependent turnover of
CC       intracellular proteins that initiate with Met-Asn. These proteins
CC       are acetylated on the retained initiator methionine and can
CC       subsequently be modified by the removal of N-acetyl methionine by
CC       acylaminoacid hydrolase (AAH). Conversion of the resulting N-
CC       terminal asparagine to aspartate by PNAD renders the protein
CC       susceptible to arginylation, polyubiquitinylation and degradation
CC       as specified by the N-end rule. This enzyme does not act on
CC       substrates with internal or C-terminal asparagines and does not
CC       act on glutamine residues in any position.
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U57692; AAB66490.1; -.
DR   EMBL; U57691; AAC52885.1; -.
DR   MGD; MGI:108471; Ntan1.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0008418; F:protein N-terminal asparagine amidohydrolas. . .; IDA.
DR   GO; GO:0008344; P:adult locomotory behavior; IMP.
DR   GO; GO:0007613; P:memory; IMP.
KW   Hydrolase.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   309 AA;  34464 MW;  BBB183EDD8F2521F CRC64;

  Query Match          77.1%;  Score 37;  DB 1;  Length 309;
```

```
  Best Local Similarity   85.7%;  Pred. No. 3.9;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSH 7
              |||||:|
Db         85 GATCLTH 91


RESULT 4
PNAD_PIG
ID   PNAD_PIG       STANDARD;      PRT;   309 AA.
AC   Q28955;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein N-terminal asparagine amidohydrolase (EC 3.5.1.-) (Protein
DE   NH2-terminal asparagine deamidase) (N-terminal Asn amidase) (NTN-
DE   amidase) (PNAD) (Protein NH2-terminal asparagine amidohydrolase)
DE   (PNAA).
GN   NTAN1.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-17; 106-118; 214-227 AND
RP   233-248.
RC   TISSUE=Liver;
RX   MEDLINE=95113832; PubMed=7814382;
RA   Stewart A.E., Arfin S.M., Bradshaw R.A.;
RT   "The sequence of porcine protein NH2-terminal asparagine
RT   amidohydrolase. A new component of the N-end Rule pathway.";
RL   J. Biol. Chem. 270:25-28(1995).
RN   [2]
RP   CHARACTERIZATION.
RC   TISSUE=Liver;
RX   MEDLINE=94375451; PubMed=8089117;
RA   Stewart A.E., Arfin S.M., Bradshaw R.A.;
RT   "Protein NH2-terminal asparagine deamidase. Isolation and
RT   characterization of a new enzyme.";
RL   J. Biol. Chem. 269:23509-23517(1994).
CC   -!- FUNCTION: Side chain deamidation of N-terminal asparagine residues
CC       to aspartate. Required for the ubiquitin-dependent turnover of
CC       intracellular proteins that initiate with Met-Asn. These proteins
CC       are acetylated on the retained initiator methionine and can
CC       subsequently be modified by the removal of N-acetyl methionine by
CC       acylaminoacid hydrolase (AAH). Conversion of the resulting N-
CC       terminal asparagine to aspartate by PNAD renders the protein
CC       susceptible to arginylation, polyubiquitinylation and degradation
CC       as specified by the N-end rule. This enzyme does not act on
CC       substrates with internal or C-terminal asparagines and does not
CC       act on glutamine residues in any position.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17062; AAA65019.1; -.
DR   PIR; A55768; A55768.
KW   Hydrolase.
FT   INIT_MET      0      0
SQ   SEQUENCE   309 AA;  34629 MW;  C527D692823CE547 CRC64;

  Query Match             77.1%;  Score 37;  DB 1;  Length 309;
  Best Local Similarity   85.7%;  Pred. No. 3.9;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSH 7
              |||||:|
Db         85 GATCLTH 91


RESULT 5
DHAY_YEAST
ID   DHAY_YEAST     STANDARD;      PRT;   511 AA.
AC   P32872;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Aldehyde dehydrogenase 2, mitochondrial precursor (EC 1.2.1.3).
GN   ALD2 OR ALDH2.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Thielen J.;
RL   Thesis (1993), Heinrich-Heine University / Duesseldorf, Germany.
CC   -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC   -!- PATHWAY: Ethanol utilization; second step.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial matrix (Potential).
CC   -!- SIMILARITY: Belongs to the aldehyde dehydrogenase family.
CC   -!- CAUTION: Is not present in yeast genome.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z17314; CAA78962.1; -.
DR   PIR; S31308; S31308.
DR   HSSP; P05091; 1CW3.
DR   SGD; L0000075; ALD2.
DR   InterPro; IPR002086; Aldehyde_dehydr.
DR   Pfam; PF00171; aldedh; 1.
DR   PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR   PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW   Oxidoreductase; NAD; Mitochondrion; Transit peptide.
FT   TRANSIT       1     21       MITOCHONDRION.
FT   CHAIN        22    511       ALDEHYDE DEHYDROGENASE 2.
FT   NP_BIND     274    279       NAD (ADP PART) (BY SIMILARITY).
FT   ACT_SITE    297    297       BY SIMILARITY.
FT   ACT_SITE    331    331       BY SIMILARITY.
SQ   SEQUENCE   511 AA;  56466 MW;  70EDAE951B84EE4A CRC64;

  Query Match             72.9%;  Score 35;  DB 1;  Length 511;
  Best Local Similarity   85.7%;  Pred. No. 17;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GATCLSH 7
              || ||||
Db        258 GAACLSH 264


RESULT 6
TNG2_HUMAN
ID   TNG2_HUMAN     STANDARD;      PRT;   110 AA.
AC   P56847;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   TNG2 protein.
GN   TNG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20056259; PubMed=10588720;
RA   Hallas C., Pekarsky Y., Itoyama T., Varnum J., Bichi R.,
RA   Rothstein J.L., Croce C.M.;
RT   "Genomic analysis of human and mouse TCL1 loci reveals a complex of
RT   tightly clustered genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:14418-14423(1999).
CC   -!- TISSUE SPECIFICITY: Expressed in many tissues.
CC   -!- DISEASE: ACTIVATED IN T CELL LEUKEMIAS WITH REARRANGEMENTS AT
CC       14Q32.1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF195821; AAF07204.1; -.
DR   MIM; 604412; -.
SQ   SEQUENCE   110 AA;  12856 MW;  D604B1C43A9C123B CRC64;

  Query Match             70.8%;  Score 34;  DB 1;  Length 110;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCLSHS 8
              ||||||
Db         88 TCLSHS 93


RESULT 7
FAS2_SCHAM
ID   FAS2_SCHAM     STANDARD;      PRT;   898 AA.
AC   P22648;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Fasciclin II precursor (FAS II).
GN   FAS2.
OS   Schistocerca americana (American grasshopper).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridomorpha;
OC   Acridoidea; Acrididae; Cyrtacanthacridinae; Schistocerca.
OX   NCBI_TaxID=7009;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89043938; PubMed=3187519;
RA   Harrelson A.L., Goodman C.S.;
RT   "Growth cone guidance in insects: fasciclin II is a member of the
RT   immunoglobulin superfamily.";
RL   Science 242:700-708(1988).
RN   [2]
RP   SEQUENCE OF 423-436.
RX   MEDLINE=88276943; PubMed=2839842;
RA   Snow P.M., Zinn K., Harrelson A.L., McAllister L., Schilling J.,
RA   Bastiani M.J., Makk G., Goodman C.S.;
RT   "Characterization and cloning of fasciclin I and fasciclin II
RT   glycoproteins in the grasshopper.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:5291-5295(1988).
CC   -!- FUNCTION: NEURONAL RECOGNITION MOLECULE. INVOLVED IN A PATHWAY
CC       RECOGNITION FOR AXONS DURING THE DEVELOPMENT OF NERVE FASCICLES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 2 fibronectin type III domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03789; AAA29810.1; -.
DR   PIR; A40114; A40114.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 5.
KW   Cell adhesion; Glycoprotein; Repeat; Immunoglobulin domain;
KW   Transmembrane; Signal; Neurogenesis.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    898       FASCICLIN II.
FT   DOMAIN       23    764       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    765    782       POTENTIAL.
FT   DOMAIN      783    898       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       31    124       IG-LIKE C2-TYPE 1.
FT   DOMAIN      134    219       IG-LIKE C2-TYPE 2.
FT   DOMAIN      226    316       IG-LIKE C2-TYPE 3.
FT   DOMAIN      321    423       IG-LIKE C2-TYPE 4.
FT   DOMAIN      428    525       IG-LIKE C2-TYPE 5.
FT   DOMAIN      548    611       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      662    730       FIBRONECTIN TYPE-III 2.
FT   DISULFID     48    113       POTENTIAL.
FT   DISULFID    156    203       POTENTIAL.
FT   DISULFID    248    300       POTENTIAL.
FT   DISULFID    343    407       POTENTIAL.
FT   DISULFID    450    509       POTENTIAL.
FT   CARBOHYD     35     35       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    149    149       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    192    192       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    297    297       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    447    447       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    457    457       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   898 AA;  99064 MW;  07989EA4F39604AC CRC64;

  Query Match             70.8%;  Score 34;  DB 1;  Length 898;
  Best Local Similarity   71.4%;  Pred. No. 50;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 TCLSHSV 9
              |||:||:
Db        449 TCLAHSI 455


RESULT 8
KNTC_HUMAN
ID   KNTC_HUMAN     STANDARD;      PRT;  2209 AA.
AC   P50748;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Kinetochore-associated protein 1 (Rough deal homolog) (hRod) (HsROD)
DE   (Rod).
GN   KNTC1 OR KIAA0166.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
```

```
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
RN   [2]
RP   FUNCTION, AND SUBCELLULAR LOCATION.
RX   MEDLINE=20582455; PubMed=11146660;
RA   Chan G.K.T., Jablonski S.A., Starr D.A., Goldberg M.L., Yen T.J.;
RT   "Human Zw10 and ROD are mitotic checkpoint proteins that bind to
RT   kinetochores.";
RL   Nat. Cell Biol. 2:944-947(2000).
RN   [3]
RP   FUNCTION, AND SUBCELLULAR LOCATION.
RX   MEDLINE=21474370; PubMed=11590237;
RA   Scaeerou F., Starr D.A., Piano F., Papoulas O., Karess R.E.,
RA   Goldberg M.L.;
RT   "The ZW10 and Rough Deal checkpoint proteins function together in a
RT   large, evolutionarily conserved complex targeted to the kinetochore.";
RL   J. Cell Sci. 114:3103-3114(2001).
CC   -!- FUNCTION: Essential component of the mitotic checkpoint, which
CC       prevents cells from prematurely exiting mitosis. Required for the
CC       assembly of the dynein-dynactin complex onto kinetochores.
CC   -!- SUBUNIT: Interacts with Zw10. This interaction is required for
CC       stable association with the kinetochore.
CC   -!- SUBCELLULAR LOCATION: Dynamic pattern of localization during the
CC       cell cycle. At interphase, uniformly distributed thoughout the
CC       cytoplasm and nucleus. By prophase and until late stages of
CC       prometaphase, a fraction of the total pool is concentrated at
CC       kinetochores. By metaphase, detected at kinetochores, along
CC       spindle fibers and most prominently at the poles. By late anaphase
CC       until the end of telophase, no longer detectable on kinetochores
CC       or along spindle fibers, but still present at the spindle poles.
CC   -!- TISSUE SPECIFICITY: High expression in testis.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D79988; BAA11483.1; -.
DR   Genew; HGNC:17255; KNTC1.
DR   GK; P50748; -.
DR   MIM; 607363; -.
DR   GO; GO:0005828; C:kinetochore microtubule; IDA.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0000922; C:spindle pole; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0007093; P:mitotic checkpoint; IDA.
DR   GO; GO:0006461; P:protein complex assembly; NAS.
DR   GO; GO:0007096; P:regulation of exit from mitosis; NAS.
KW   Mitosis; Cell cycle; Nuclear protein.
SQ   SEQUENCE   2209 AA;  250746 MW;  514C948733CBFDAC CRC64;

  Query Match             70.8%;  Score 34;  DB 1;  Length 2209;
  Best Local Similarity   85.7%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 TCLSHSV 9
              ||| |||
Db       1281 TCLQHSV 1287


RESULT 9
DEF_AGRT5
ID   DEF_AGRT5      STANDARD;      PRT;   170 AA.
AC   Q8UID1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Peptide deformylase (EC 3.5.1.88) (PDF) (Polypeptide deformylase).
GN   DEF OR ATU0366 OR AGR_C_640.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
CC   -!- FUNCTION: Removes the formyl group from the N-terminal Met of
CC       newly synthesized proteins. Requires at least a dipeptide for an
CC       efficient rate of reaction. N-terminal L-methionine is a
CC       prerequisite for activity but the enzyme has broad specificity at
CC       other positions (By similarity).
CC   -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC       methionyl peptide.
CC   -!- COFACTOR: Binds 1 iron(II) ion (By similarity).
CC   -!- SIMILARITY: Belongs to the polypeptide deformylase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE009008; AAL41388.1; -.
DR   EMBL; AE007974; AAK86183.1; -.
DR   PIR; AF2621; AF2621.
DR   PIR; F97403; F97403.
DR   HAMAP; MF_00163; -; 1.
DR   InterPro; IPR000181; Pep_deformylase.
DR   Pfam; PF01327; Pep_deformylase; 1.
DR   PRINTS; PR01576; PDEFORMYLASE.
DR   ProDom; PD003844; Pep_deformylase; 1.
DR   TIGRFAMs; TIGR00079; pept_deformyl; 1.
KW   Protein biosynthesis; Hydrolase; Iron; Complete proteome.
FT   ACT_SITE    137    137       BY SIMILARITY.
FT   METAL        94     94       IRON (BY SIMILARITY).
FT   METAL       136    136       IRON (BY SIMILARITY).
FT   METAL       140    140       IRON (BY SIMILARITY).
SQ   SEQUENCE   170 AA;  19026 MW;  CAA0DE1D6497AD22 CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 170;
  Best Local Similarity   62.5%;  Pred. No. 13;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 ATCLSHSV 9
```

```
              |||| | :
Db        131 ATCLQHEI 138


RESULT 10
DEF_RHIME
ID   DEF_RHIME      STANDARD;      PRT;   174 AA.
AC   Q92SH6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Peptide deformylase (EC 3.5.1.88) (PDF) (Polypeptide deformylase).
GN   DEF OR R00419 OR SMC01101.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC   -!- FUNCTION: Removes the formyl group from the N-terminal Met of
CC       newly synthesized proteins. Requires at least a dipeptide for an
CC       efficient rate of reaction. N-terminal L-methionine is a
CC       prerequisite for activity but the enzyme has broad specificity at
CC       other positions (By similarity).
CC   -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC       methionyl peptide.
CC   -!- COFACTOR: Binds 1 iron(II) ion (By similarity).
CC   -!- SIMILARITY: Belongs to the polypeptide deformylase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL591783; CAC41856.1; ALT_INIT.
DR   HAMAP; MF_00163; -; 1.
DR   InterPro; IPR000181; Pep_deformylase.
DR   Pfam; PF01327; Pep_deformylase; 1.
DR   PRINTS; PR01576; PDEFORMYLASE.
DR   ProDom; PD003844; Pep_deformylase; 1.
DR   TIGRFAMs; TIGR00079; pept_deformyl; 1.
KW   Protein biosynthesis; Hydrolase; Iron; Complete proteome.
FT   ACT_SITE    137    137       BY SIMILARITY.
FT   METAL        94     94       IRON (BY SIMILARITY).
FT   METAL       136    136       IRON (BY SIMILARITY).
FT   METAL       140    140       IRON (BY SIMILARITY).
SQ   SEQUENCE   174 AA;  19476 MW;  9782A6304F12D6D8 CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 174;
  Best Local Similarity   62.5%;  Pred. No. 13;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 ATCLSHSV 9
              |||| | :
Db        131 ATCLQHEI 138


RESULT 11
DEF_BRUME
ID   DEF_BRUME      STANDARD;      PRT;   175 AA.
AC   Q8YDB4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Peptide deformylase (EC 3.5.1.88) (PDF) (Polypeptide deformylase).
GN   DEF OR BMEII0264 OR BRA1035.
OS   Brucella melitensis, and
OS   Brucella suis.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29459, 29461;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=B.melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
RX   MEDLINE=20020109; PubMed=11756688;
RA   DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA   Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA   Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA   Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA   Haselkorn R., Kyrpides N., Overbeek R.;
RT   "The genome sequence of the facultative intracellular pathogen
RT   Brucella melitensis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=B.suis; STRAIN=1330 / Biovar 1;
RX   MEDLINE=22247741; PubMed=12271122;
RA   Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA   Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA   Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA   Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA   Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA   Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT   "The Brucella suis genome reveals fundamental similarities between
RT   animal and plant pathogens and symbionts.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
CC   -!- FUNCTION: Removes the formyl group from the N-terminal Met of
CC       newly synthesized proteins. Requires at least a dipeptide for an
CC       efficient rate of reaction. N-terminal L-methionine is a
CC       prerequisite for activity but the enzyme has broad specificity at
CC       other positions (By similarity).
CC   -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC       methionyl peptide.
CC   -!- COFACTOR: Binds 1 iron(II) ion (By similarity).
CC   -!- SIMILARITY: Belongs to the polypeptide deformylase family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE009665; AAL53505.1; ALT_INIT.
DR   EMBL; AE014596; AAN34202.1; ALT_INIT.
DR   TIGR; BRA1035; -.
DR   HAMAP; MF_00163; -; 1.
DR   InterPro; IPR000181; Pep_deformylase.
DR   Pfam; PF01327; Pep_deformylase; 1.
DR   PRINTS; PR01576; PDEFORMYLASE.
DR   ProDom; PD003844; Pep_deformylase; 1.
DR   TIGRFAMs; TIGR00079; pept_deformyl; 1.
KW   Protein biosynthesis; Hydrolase; Iron; Complete proteome.
FT   ACT_SITE    137    137       BY SIMILARITY.
FT   METAL        94     94       IRON (BY SIMILARITY).
FT   METAL       136    136       IRON (BY SIMILARITY).
FT   METAL       140    140       IRON (BY SIMILARITY).
SQ   SEQUENCE   175 AA;  19660 MW;  873A6ABCF3EFFCE2 CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 175;
  Best Local Similarity   62.5%;  Pred. No. 14;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy          2 ATCLSHSV 9
              |||| | :
Db        131 ATCLQHEI 138


RESULT 12
DEF_RHILO
ID   DEF_RHILO      STANDARD;      PRT;   176 AA.
AC   Q98D52;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Peptide deformylase (EC 3.5.1.88) (PDF) (Polypeptide deformylase).
GN   DEF OR MLL4855.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
CC   -!- FUNCTION: Removes the formyl group from the N-terminal Met of
CC       newly synthesized proteins. Requires at least a dipeptide for an
CC       efficient rate of reaction. N-terminal L-methionine is a
CC       prerequisite for activity but the enzyme has broad specificity at
CC       other positions (By similarity).
CC   -!- CATALYTIC ACTIVITY: Formyl-L-methionyl peptide + H(2)O = formate +
CC       methionyl peptide.
CC   -!- COFACTOR: Binds 1 iron(II) ion (By similarity).
CC   -!- SIMILARITY: Belongs to the polypeptide deformylase family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP003005; BAB51419.1; -.
DR   HAMAP; MF_00163; -; 1.
DR   InterPro; IPR000181; Pep_deformylase.
DR   Pfam; PF01327; Pep_deformylase; 1.
DR   PRINTS; PR01576; PDEFORMYLASE.
DR   ProDom; PD003844; Pep_deformylase; 1.
DR   TIGRFAMs; TIGR00079; pept_deformyl; 1.
KW   Protein biosynthesis; Hydrolase; Iron; Complete proteome.
FT   ACT_SITE    137    137       BY SIMILARITY.
FT   METAL        94     94       IRON (BY SIMILARITY).
FT   METAL       136    136       IRON (BY SIMILARITY).
FT   METAL       140    140       IRON (BY SIMILARITY).
SQ   SEQUENCE   176 AA;  19651 MW;  8897DC6FBE68C045 CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 176;
  Best Local Similarity   62.5%;  Pred. No. 14;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 ATCLSHSV 9
              |||| | :
Db        131 ATCLQHEI 138


RESULT 13
```

```
DLK_HUMAN
ID   DLK_HUMAN      STANDARD;      PRT;   383 AA.
AC   P80370; P15803;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Delta-like protein precursor (DLK) (pG2) [Contains: Fetal antigen 1
DE   (FA1)].
GN   DLK1 OR DLK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=93179372; PubMed=8095043;
RA   Laborda J., Sausville E.A., Hoffman T., Notario V.;
RT   "dlk, a putative mammalian homeotic gene differentially expressed in
RT   small cell lung carcinoma and neuroendocrine tumor cell line.";
RL   J. Biol. Chem. 268:3817-3820(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland, and Placenta;
RX   MEDLINE=95226449; PubMed=7711066;
RA   Lee Y.L., Helman L., Hoffman T., Laborda J.;
RT   "dlk, pG2 and Pref-1 mRNAs encode similar proteins belonging to the
RT   EGF-like superfamily. Identification of polymorphic variants of this
RT   RNA.";
RL   Biochim. Biophys. Acta 1261:223-232(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=90175023; PubMed=2308864;
RA   Helman L.J., Sack N., Plon S., Israel M.A.;
RT   "The sequence of an adrenal specific human cDNA, pG2.";
RL   Nucleic Acids Res. 18:685-685(1990).
RN   [4]
RP   SEQUENCE OF 24-383.
RC   TISSUE=Amniotic fluid;
RX   MEDLINE=95010145; PubMed=7925474;
RA   Jensen C.H., Krogh T.N., Hoejrup P., Clausen P.P., Skjoedt K.,
RA   Larsson L.-I., Enghild J.J., Teisner B.;
RT   "Protein structure of fetal antigen 1 (FA1). A novel circulating
RT   human epidermal-growth-factor-like protein expressed in
RT   neuroendocrine tumors and its relation to the gene products of dlk
RT   and pG2.";
RL   Eur. J. Biochem. 225:83-92(1994).
RN   [5]
RP   SEQUENCE OF 24-60.
RC   TISSUE=Amniotic fluid;
RX   MEDLINE=93273893; PubMed=8501199;
RA   Jensen C.H., Teisner B., Hoejrup P., Rasmussen H.B., Madsen O.D.,
RA   Nielsen B., Skjoedt K.;
RT   "Studies on the isolation, structural analysis and tissue localization
RT   of fetal antigen 1 and its relation to a human adrenal-specific cDNA,
RT   pG2.";
RL   Hum. Reprod. 8:635-641(1993).
CC   -!- FUNCTION: May have a role in neuroendocrine differentiation.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=Long;
CC         IsoId=P80370-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=P80370-2; Sequence=VSP_001377;
CC   -!- TISSUE SPECIFICITY: Found within the stromal cells in close
CC       contact to the vascular structure of placental villi, yolk sac,
CC       fetal liver, adrenal cortex and pancreas and in the beta cells of
CC       the islets of langerhans in the adult pancreas. Found also in some
CC       forms of neuroendocrine lung tumor tissue.
```

```
CC   -!- PTM: N- and O-glycosylated.
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC       in positions 23, 45, 108, 199.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z12172; CAA78163.1; -.
DR   EMBL; U15979; AAA75364.1; -.
DR   EMBL; U15981; AAA75365.1; -.
DR   EMBL; X17544; CAA35582.1; ALT_FRAME.
DR   PIR; S53716; S53716.
DR   HSSP; P08709; 1BF9.
DR   Genew; HGNC:2907; DLK1.
DR   MIM; 176290; -.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0005803; C:secretory vesicle; TAS.
DR   GO; GO:0005625; C:soluble fraction; TAS.
DR   GO; GO:0007275; P:development; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   Pfam; PF00008; EGF; 5.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00181; EGF; 6.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS50026; EGF_3; 5.
KW   Antigen; Transmembrane; EGF-like domain; Repeat; Glycoprotein; Signal;
KW   Alternative splicing.
FT   SIGNAL       1     23
FT   CHAIN       24    383       DELTA-LIKE PROTEIN.
FT   CHAIN       24    303       FETAL ANTIGEN 1.
FT   DOMAIN      24    303       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   304    327       POTENTIAL.
FT   DOMAIN     328    383       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      24     55       EGF-LIKE 1.
FT   DOMAIN      53     86       EGF-LIKE 2.
FT   DOMAIN      88    125       EGF-LIKE 3.
FT   DOMAIN     127    168       EGF-LIKE 4.
FT   DOMAIN     170    206       EGF-LIKE 5.
FT   DOMAIN     208    245       EGF-LIKE 6.
FT   DISULFID    26     37       BY SIMILARITY.
FT   DISULFID    30     43       BY SIMILARITY.
FT   DISULFID    45     54       BY SIMILARITY.
FT   DISULFID    57     68       BY SIMILARITY.
FT   DISULFID    63     74       BY SIMILARITY.
FT   DISULFID    76     85       BY SIMILARITY.
FT   DISULFID    92    103       BY SIMILARITY.
FT   DISULFID    97    113       BY SIMILARITY.
FT   DISULFID   115    124       BY SIMILARITY.
FT   DISULFID   131    144       BY SIMILARITY.
FT   DISULFID   138    156       BY SIMILARITY.
FT   DISULFID   158    167       BY SIMILARITY.
FT   DISULFID   174    185       BY SIMILARITY.
FT   DISULFID   179    194       BY SIMILARITY.
FT   DISULFID   196    205       BY SIMILARITY.
FT   DISULFID   212    223       BY SIMILARITY.
FT   DISULFID   217    233       BY SIMILARITY.
FT   DISULFID   235    244       BY SIMILARITY.
FT   CARBOHYD    94     94       O-LINKED.
FT   CARBOHYD   100    100       N-LINKED (GLCNAC. . .).
FT   CARBOHYD   143    143       O-LINKED.
FT   CARBOHYD   163    163       O-LINKED (PARTIAL).
FT   CARBOHYD   165    165       N-LINKED (GLCNAC. . .) (PARTIAL).
FT   CARBOHYD   172    172       N-LINKED (GLCNAC. . .) (PARTIAL).
FT   CARBOHYD   214    214       O-LINKED.
FT   CARBOHYD   222    222       O-LINKED (PARTIAL).
FT   CARBOHYD   251    251       O-LINKED (PARTIAL).
FT   CARBOHYD   260    260       O-LINKED (PARTIAL).
FT   VARSPLIC   229    301       Missing (in isoform Short).
FT                               /FTId=VSP_001377.
FT   VARIANT    347    347       Missing (in clone HDLKAAG).
FT                               /FTId=VAR_002274.
FT   CONFLICT    45     47       MISSING (IN REF. 3).
FT   CONFLICT    46     47       QP -> HV (IN REF. 1).
FT   CONFLICT   108    108       G -> D (IN REF. 4).
SQ   SEQUENCE   383 AA;  41143 MW;  2C55C6E355B4415B CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 383;
  Best Local Similarity   62.5%;  Pred. No. 32;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GATCLSHS 8
              | ||| |:
Db        220 GGTCLQHT 227


RESULT 14
DLK_MOUSE
ID   DLK_MOUSE      STANDARD;      PRT;   385 AA.
AC   Q09163; Q07645; Q62208;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Delta-like protein precursor (DLK) (Preadipocyte factor 1) (Pref-1)
DE   (Adipocyte differentiation inhibitor protein) [Contains: Fetal antigen
DE   1 (FA1)].
GN   DLK1 OR DLK OR PREF1 OR SCP-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss; TISSUE=Fibroblast;
RX   MEDLINE=93179372; PubMed=8095043;
RA   Laborda J., Sausville E.A., Hoffman T., Notario V.;
RT   "dlk, a putative mammalian homeotic gene differentially expressed in
RT   small cell lung carcinoma and neuroendocrine tumor cell line.";
RL   J. Biol. Chem. 268:3817-3820(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93272313; PubMed=8500166;
RA   Smas C.M., Sul H.S.;
RT   "Pref-1, a protein containing EGF-like repeats, inhibits adipocyte
RT   differentiation.";
RL   Cell 73:725-734(1993).
RN   [3]
RP   SEQUENCE FROM N.A., AND VARIANT LYS-347 DEL.
RC   TISSUE=Adrenal gland, and Placenta;
RX   MEDLINE=95226449; PubMed=7711066;
RA   Lee Y.L., Helman L., Hoffman T., Laborda J.;
RT   "dlk, pG2 and Pref-1 mRNAs encode similar proteins belonging to the
RT   EGF-like superfamily. Identification of polymorphic variants of this
RT   RNA.";
RL   Biochim. Biophys. Acta 1261:223-232(1995).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Maruyama K., Nishijima S., Kuromitsu S., Ichikawa A., Masuda E.,
RA   Takemoto T., Kodama H., Kawashima H.;
RL   Submitted (AUG-1993) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1-8 FROM N.A., AND ALTERNATIVE SPLICING.
```

```
RX   MEDLINE=94325292; PubMed=7519443;
RA   Smas C.M., Green D., Sul H.S.;
RT   "Structural characterization and alternate splicing of the gene
RT   encoding the preadipocyte EGF-like protein pref-1.";
RL   Biochemistry 33:9257-9265(1994).
RN   [6]
RP   CARBOHYDRATE-LINKAGE SITES.
RX   MEDLINE=97234561; PubMed=9118998;
RA   Krogh T.N., Bachmann E., Teisner B., Skjoedt K., Hoejrup P.;
RT   "Glycosylation analysis and protein structure determination of murine
RT   fetal antigen 1 (mFA1) -- the circulating gene product of the
RT   delta-like protein (dlk), preadipocyte factor 1 (Pref-1) and
RT   stromal-cell-derived protein 1 (SCP-1) cDNAs.";
RL   Eur. J. Biochem. 244:334-342(1997).
CC   -!- FUNCTION: May have a role in neuroendocrine differentiation.
CC       Inhibits adipocyte differentiation.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=6;
CC         Comment=Additional isoforms seem to exist;
CC       Name=A;
CC         IsoId=Q09163-1; Sequence=Displayed;
CC       Name=B;
CC         IsoId=Q09163-2; Sequence=VSP_001380;
CC       Name=C;
CC         IsoId=Q09163-3; Sequence=VSP_001381;
CC       Name=C2;
CC         IsoId=Q09163-4; Sequence=VSP_001382;
CC       Name=D;
CC         IsoId=Q09163-5; Sequence=VSP_001378;
CC       Name=D2;
CC         IsoId=Q09163-6; Sequence=VSP_001379;
CC   -!- PTM: N- and O-glycosylated.
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z12171; CAA78162.1; -.
DR   EMBL; U15980; AAB60495.1; -.
DR   EMBL; L12721; AAA37175.1; -.
DR   EMBL; S71340; -; NOT_ANNOTATED_CDS.
DR   EMBL; D16847; BAA04121.1; -.
DR   PIR; A54785; A54785.
DR   PIR; S53718; S53718.
DR   HSSP; P00740; 1EDM.
DR   GlycoSuiteDB; Q09163; -.
DR   MGD; MGI:94900; Dlk1.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   Pfam; PF00008; EGF; 5.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50026; EGF_3; 6.
KW   Antigen; Transmembrane; EGF-like domain; Repeat; Glycoprotein; Signal;
KW   Alternative splicing.
FT   SIGNAL       1     23       BY SIMILARITY.
FT   CHAIN        24    385      DELTA-LIKE PROTEIN.
FT   CHAIN        24    305      FETAL ANTIGEN 1.
FT   DOMAIN       24    305      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     306   329      POTENTIAL.
FT   DOMAIN       330   385      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       24    55       EGF-LIKE 1.
FT   DOMAIN       53    86       EGF-LIKE 2.
FT   DOMAIN       88    125      EGF-LIKE 3.
FT   DOMAIN       127   168      EGF-LIKE 4.
FT   DOMAIN       172   208      EGF-LIKE 5.
FT   DOMAIN       210   247      EGF-LIKE 6.
FT   DISULFID     26    37       BY SIMILARITY.
FT   DISULFID     30    43       BY SIMILARITY.
FT   DISULFID     45    54       BY SIMILARITY.
FT   DISULFID     57    68       BY SIMILARITY.
FT   DISULFID     63    74       BY SIMILARITY.
FT   DISULFID     76    85       BY SIMILARITY.
FT   DISULFID     92    103      BY SIMILARITY.
FT   DISULFID     97    113      BY SIMILARITY.
FT   DISULFID     115   124      BY SIMILARITY.
FT   DISULFID     131   144      BY SIMILARITY.
FT   DISULFID     138   156      BY SIMILARITY.
FT   DISULFID     158   167      BY SIMILARITY.
FT   DISULFID     176   187      BY SIMILARITY.
FT   DISULFID     181   196      BY SIMILARITY.
FT   DISULFID     198   207      BY SIMILARITY.
FT   DISULFID     214   225      BY SIMILARITY.
FT   DISULFID     219   235      BY SIMILARITY.
FT   DISULFID     237   246      BY SIMILARITY.
FT   CARBOHYD     94    94       O-LINKED.
FT                               /FTId=CAR_000160.
FT   CARBOHYD     100   100      N-LINKED (GLCNAC. . .).
FT                               /FTId=CAR_000183.
FT   CARBOHYD     165   165      N-LINKED (GLCNAC. . .).
FT   CARBOHYD     174   174      N-LINKED (GLCNAC. . .).
FT   CARBOHYD     216   216      O-LINKED.
FT                               /FTId=CAR_000161.
FT   CARBOHYD     224   224      O-LINKED.
FT                               /FTId=CAR_000162.
FT   CARBOHYD     258   258      O-LINKED.
FT   CARBOHYD     267   267      O-LINKED (PARTIAL).
FT   CARBOHYD     271   271      O-LINKED.
FT   CARBOHYD     295   295      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC     211   303      Missing (in isoform D).
FT                               /FTId=VSP_001378.
FT   VARSPLIC     211   305      Missing (in isoform D2).
FT                               /FTId=VSP_001379.
FT   VARSPLIC     231   281      Missing (in isoform B).
FT                               /FTId=VSP_001380.
FT   VARSPLIC     231   303      Missing (in isoform C).
FT                               /FTId=VSP_001381.
FT   VARSPLIC     231   305      Missing (in isoform C2).
FT                               /FTId=VSP_001382.
FT   VARIANT      347   347      MISSING.
FT   CONFLICT     250   250      R -> P (IN REF. 2).
FT   CONFLICT     320   385      VLGTVAIVFLNKCETWVSNLRYNHTFRKKKNLLLQYNSGEE
FT                               LAVNIIFPEKIDMTTFNKEAGDEEI -> CWAPWPSSFSTS
FT                               AKPGCPTCATTTCFARRRTSCCSITAARSWRSISSSPRRLT
FT                               (IN REF. 2).
FT   CONFLICT     344   345      TF -> ML (IN REF. 4).
SQ   SEQUENCE   385 AA;  41320 MW;  E79864FEA5AF4FF1 CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 385;
  Best Local Similarity   62.5%;  Pred. No. 32;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GATCLSHS 8
              | ||| |:
Db        222 GGTCLQHT 229


RESULT 15
YEY2_YEAST
ID   YEY2_YEAST     STANDARD;      PRT;   443 AA.
AC   P10356;
```

```
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical 49.5 kDa protein in UBP3-PET122 intergenic region.
GN   YER152C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313264; PubMed=9169868;
RA   Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA   Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA   Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA   Hunicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA   Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA   Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA   Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL   Nature 387:78-81(1997).
RN   [2]
RP   SEQUENCE OF 1-230 FROM N.A.
RC   STRAIN=AB320;
RX   MEDLINE=89083497; PubMed=2849752;
RA   Ohmen J.D., Kloeckener-Gruissem B., McEwen J.E.;
RT   "Molecular cloning and nucleotide sequence of the nuclear PET122 gene
RT   required for expression of the mitochondrial COX3 gene in S.
RT   cerevisiae.";
RL   Nucleic Acids Res. 16:10783-10802(1988).
RN   [3]
RP   REVISIONS, SEQUENCE OF 1-230 FROM N.A.
RX   MEDLINE=90258894; PubMed=2160592;
RA   Ohmen J.D., Burke K.A., McEwen J.E.;
RT   "Divergent overlapping transcripts at the PET122 locus in
RT   Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 10:3027-3035(1990).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U18917; AAB64679.1; -.
DR   EMBL; X07558; CAA30440.1; -.
DR   PIR; S50655; QQBYPT.
DR   GermOnline; 139230; -.
DR   SGD; S0000954; YER152C.
KW   Hypothetical protein.
SQ   SEQUENCE   443 AA;  49490 MW;  BCA67A4D3B9D7A14 CRC64;

  Query Match             68.8%;  Score 33;  DB 1;  Length 443;
  Best Local Similarity   55.6%;  Pred. No. 37;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              || | ||::
Db        429 GAVCKSHAI 437


Search completed: August 25, 2004, 02:27:45
Job time : 2.48043 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 25, 2004, 02:22:38 ; Search time 2.43416 Seconds
(without alignments)
1166.587 Million cell updates/sec

Title: US-09-593-793A-113_COPY_367_375
Perfect score: 48
Sequence: 1 GATCLSHSV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	48	100.0	501	6	Q95KC5	Q95kc5 macaca fasc
2	48	100.0	553	4	Q96JT2	Q96jt2 homo sapien
3	48	100.0	553	6	Q95KI5	Q95ki5 macaca fasc
4	42	87.5	450	11	Q8K252	Q8k252 mus musculu
5	42	87.5	553	11	Q8K0H7	Q8k0h7 mus musculu
6	39	81.2	1406	4	Q8WWY0	Q8wwy0 homo sapien
7	38	79.2	157	13	Q98882	Q98882 brachydanio
8	38	79.2	162	13	Q9IAK7	Q9iak7 brachydanio
9	38	79.2	162	13	Q9IAK6	Q9iak6 brachydanio
10	37	77.1	162	13	Q9IAK5	Q9iak5 brachydanio
11	37	77.1	293	11	Q8BRX4	Q8brx4 mus musculu
12	37	77.1	310	4	Q7Z4Z0	Q7z4z0 homo sapien
13	37	77.1	669	5	O44143	O44143 caenorhabdi
14	36	75.0	135	6	Q9GL46	Q9gl46 sus scrofa
15	36	75.0	308	6	O46370	O46370 bos taurus
16	36	75.0	687	10	Q94H65	Q94h65 oryza sativ
17	36	75.0	720	13	Q7ZYM2	Q7zym2 xenopus lae
18	36	75.0	1208	5	Q23133	Q23133 caenorhabdi
19	35	72.9	251	2	Q93A69	Q93a69 gamma-prote
20	35	72.9	266	16	Q8EBW3	Q8ebw3 shewanella
21	35	72.9	351	2	Q7X4Y4	Q7x4y4 rhodospiril
22	35	72.9	662	11	Q8VE71	Q8ve71 mus musculu
23	35	72.9	674	5	Q8T4N9	Q8t4n9 strongyloce
24	35	72.9	1082	11	Q8K0I6	Q8k0i6 mus musculu
25	35	72.9	1383	11	Q7TT36	Q7tt36 mus musculu
26	35	72.9	1479	5	Q9BKY3	Q9bky3 leishmania
27	34	70.8	105	4	Q9NPH1	Q9nph1 homo sapien
28	34	70.8	119	4	Q9P2S1	Q9p2s1 homo sapien
29	34	70.8	140	16	Q8P9P9	Q8p9p9 xanthomonas
30	34	70.8	156	3	Q12479	Q12479 saccharomyc
31	34	70.8	163	4	Q9P2S0	Q9p2s0 homo sapien
32	34	70.8	235	17	Q976P7	Q976p7 sulfolobus
33	34	70.8	267	16	Q81PP5	Q81pp5 bacillus an
34	34	70.8	267	16	Q81CJ7	Q81cj7 bacillus ce
35	34	70.8	270	5	Q9N364	Q9n364 caenorhabdi
36	34	70.8	311	2	P97031	P97031 bacillus su
37	34	70.8	311	16	O31589	O31589 bacillus su
38	34	70.8	349	16	Q8DI60	Q8di60 synechococc
39	34	70.8	408	4	Q9UGL4	Q9ugl4 homo sapien
40	34	70.8	425	4	Q8NAH1	Q8nah1 homo sapien
41	34	70.8	545	16	Q7VPB0	Q7vpb0 haemophilus
42	34	70.8	547	5	Q8WRC8	Q8wrc8 plasmodium
43	34	70.8	911	13	Q7ZZT0	Q7zzt0 brachydanio
44	34	70.8	1074	4	Q86T46	Q86t46 homo sapien
45	34	70.8	1087	4	Q86T55	Q86t55 homo sapien

ALIGNMENTS

RESULT 1
Q95KC5
ID Q95KC5 PRELIMINARY; PRT; 501 AA.
AC Q95KC5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Medulla oblongata;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB062977; BAB60745.1; -.
KW Hypothetical protein.
SQ SEQUENCE 501 AA; 53447 MW; 8C554BBD04EE0470 CRC64;

Query Match 100.0%; Score 48; DB 6; Length 501;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCLSHSV 9
||||||||| |
Db 315 GATCLSHSV 323

RESULT 2
Q96JT2
ID Q96JT2 PRELIMINARY; PRT; 553 AA.
AC Q96JT2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Prostein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Prostate;
RX   MEDLINE=21139094; PubMed=11245466;
RA   Xu J., Kalos M., Stolk J.A., Zasloff E.J., Zhang X., Houghton R.L.,
RA   Filho A.M., Nolasco M., Badaro R., Reed S.G.;
RT   "Identification and characterization of prostein, a novel prostate-
RT   specific protein.";
RL   Cancer Res. 61:1563-1568(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral Nervous System;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral Nervous System;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY033593; AAK54386.1; -.
DR   EMBL; BC050416; AAH50416.1; -.
SQ   SEQUENCE   553 AA;  59322 MW;  0AFA23FBC742A667 CRC64;

  Query Match             100.0%;  Score 48;  DB 4;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              |||||||||
Db        367 GATCLSHSV 375


RESULT 3
Q95KI5
ID   Q95KI5      PRELIMINARY;      PRT;   553 AA.
AC   Q95KI5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Temporal cortex;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB060851; BAB46871.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   553 AA;  59392 MW;  0718F3A91FB3BF1E CRC64;

  Query Match             100.0%;  Score 48;  DB 6;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 0.16;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              |||||||||
Db        367 GATCLSHSV 375


RESULT 4
Q8K252
ID   Q8K252      PRELIMINARY;      PRT;   450 AA.
AC   Q8K252;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   PROSTEIN OR 2210413P12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC034084; AAH34084.1; -.
DR   MGD; MGI:1922082; Prostein.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
DR   GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR005805; Rieske.
DR   PROSITE; PS00200; RIESKE_2; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   450 AA;  48261 MW;  72C703AEC131302C CRC64;

  Query Match             87.5%;  Score 42;  DB 11;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 2.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 ATCLSHSV 9
              ||||||||
Db        265 ATCLSHSV 272


RESULT 5
Q8K0H7
ID   Q8K0H7      PRELIMINARY;      PRT;   553 AA.
AC   Q8K0H7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (PROSTEIN homolog).
GN   PROSTEIN OR 2210413P12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
```

```
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Urinary bladder;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; BC031381; AAH31381.1; -.
DR   EMBL; AK035428; BAC29063.1; -.
DR   MGD; MGI:1922082; Prostein.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
DR   GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR005805; Rieske.
DR   PROSITE; PS00200; RIESKE_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   553 AA;  59742 MW;  6A474FEA630B4B13 CRC64;

  Query Match             87.5%;  Score 42;  DB 11;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 ATCLSHSV 9
              ||||||||
Db        368 ATCLSHSV 375


RESULT 6
Q8WWY0
ID   Q8WWY0      PRELIMINARY;      PRT;  1406 AA.
AC   Q8WWY0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CRB1 isoform II precursor.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99438399; PubMed=10508521;
RA   den Hollander A.I., ten Brink J.B., de Kok Y.J.M., van Soest S.,
RA   van den Born I.L., van Driel M.A., van de Pol D.J.R., Payne A.M.,
RA   Bhattacharya S.S., Kellner U., Hoyng C.B., Westerveld A.,
RA   Brunner H.G., Bleeker-Wagemakers E.M., Deutman A.F.,
RA   Heckenlively J.R., Cremers F.P.M., Bergen A.A.B.;
RT   "Mutations in a human homologue of Drosophila crumbs cause retinitis
RT   pigmentosa (RP12).";
RL   Nat. Genet. 23:217-221(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21592405; PubMed=11734541;
RA   den Hollander A.I., Johnson K., de Kok Y.J.M., Klebes A.,
RA   Brunner H.G., Knust E., Cremers F.P.M.;
RT   "CRB1 has a cytoplasmic domain that is functionally conserved between
RT   human and Drosophila.";
RL   Hum. Mol. Genet. 10:2767-2773(2001).
DR   EMBL; AY043325; AAL10682.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF00008; EGF; 16.
DR   Pfam; PF00054; laminin_G; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00179; EGF_CA; 8.
DR   SMART; SM00282; LamG; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 10.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS00022; EGF_1; 15.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS01187; EGF_CA; 7.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
KW   EGF-like domain; Signal.
FT   SIGNAL       5      25       POTENTIAL.
SQ   SEQUENCE   1406 AA;  154182 MW;  F2D04D20FAA6E37D CRC64;

  Query Match             81.2%;  Score 39;  DB 4;  Length 1406;
  Best Local Similarity   75.0%;  Pred. No. 29;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSHS 8
              ||||:||:
Db       1226 GATCISHT 1233


RESULT 7
Q98882
ID   Q98882      PRELIMINARY;      PRT;  157 AA.
AC   Q98882;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Odorant receptor 1 (Fragment).
GN   OR11.1 OR ZOR1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97075163; PubMed=8917589;
RA   Weth F., Nadler W., Korsching S.;
RT   "Nested expression domains for odorant receptors in zebrafish
RT   olfactory epithelium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13321-13326(1996).
DR   EMBL; U72683; AAB38866.1; -.
DR   ZFIN; ZDB-GENE-990415-190; or11.1.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     157    157
SQ   SEQUENCE   157 AA;  17656 MW;  CF5738C349E526C4 CRC64;

  Query Match             79.2%;  Score 38;  DB 13;  Length 157;
  Best Local Similarity   66.7%;  Pred. No. 5.7;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              ||||:|| :
Db        110 GATCVSHLI 118


RESULT 8
Q9IAK7
ID   Q9IAK7      PRELIMINARY;      PRT;  162 AA.
AC   Q9IAK7;
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Olfactory receptor (Fragment).
GN   OR11.1 OR DRE1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20183981; PubMed=10706615;
RA   Rouquier S., Blancher A., Giorgi D.;
RT   "The olfactory receptor gene repertoire in primates and mouse:
RT   Evidence for reduction of the functional fraction in primates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:2870-2874(2000).
DR   EMBL; AF179844; AAF40409.1; -.
DR   ZFIN; ZDB-GENE-990415-190; or11.1.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     162    162
SQ   SEQUENCE   162 AA;  18156 MW;  FA77B0A46EFCB7A9 CRC64;

  Query Match           79.2%;  Score 38;  DB 13;  Length 162;
  Best Local Similarity   66.7%;  Pred. No. 5.9;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              ||||:|| :
Db        114 GATCVSHLI 122


RESULT 9
Q9IAK6
ID   Q9IAK6       PRELIMINARY;      PRT;   162 AA.
AC   Q9IAK6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Olfactory receptor (Fragment).
GN   OR11.1 OR DRE3.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20183981; PubMed=10706615;
RA   Rouquier S., Blancher A., Giorgi D.;
RT   "The olfactory receptor gene repertoire in primates and mouse:
RT   Evidence for reduction of the functional fraction in primates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:2870-2874(2000).
DR   EMBL; AF179845; AAF40410.1; -.
DR   ZFIN; ZDB-GENE-990415-190; or11.1.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     162    162
SQ   SEQUENCE   162 AA;  18122 MW;  7877B0A46EF2B949 CRC64;

  Query Match           79.2%;  Score 38;  DB 13;  Length 162;
  Best Local Similarity   66.7%;  Pred. No. 5.9;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              ||||:|| :
Db        114 GATCVSHLI 122


RESULT 10
Q9IAK5
ID   Q9IAK5       PRELIMINARY;      PRT;   162 AA.
AC   Q9IAK5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Olfactory receptor (Fragment).
GN   OR11.1 OR DRE9.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20183981; PubMed=10706615;
RA   Rouquier S., Blancher A., Giorgi D.;
RT   "The olfactory receptor gene repertoire in primates and mouse:
RT   Evidence for reduction of the functional fraction in primates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:2870-2874(2000).
DR   EMBL; AF179846; AAF40411.1; -.
DR   ZFIN; ZDB-GENE-990415-190; or11.1.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Receptor.
FT   NON_TER       1      1
FT   NON_TER     162    162
SQ   SEQUENCE   162 AA;  18130 MW;  E1FD981D7D448C39 CRC64;

  Query Match           77.1%;  Score 37;  DB 13;  Length 162;
  Best Local Similarity   85.7%;  Pred. No. 9.5;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSH 7
              ||||:||
Db        114 GATCVSH 120


RESULT 11
Q8BRX4
ID   Q8BRX4       PRELIMINARY;      PRT;   293 AA.
AC   Q8BRX4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   N-terminal Asn amidase (Fragment).
GN   NTAN1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Aorta, and Vein;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
```

```
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK041112; BAC30825.1; -.
DR   MGD; MGI:108471; Ntan1.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0008418; F:protein N-terminal asparagine amidohydrolas. . .; IDA.
DR   GO; GO:0008344; P:adult locomotory behavior; IMP.
DR   GO; GO:0007613; P:memory; IMP.
FT   NON_TER     293    293
SQ   SEQUENCE   293 AA;  32710 MW;  8FE4E7DA8E561661 CRC64;

  Query Match             77.1%;  Score 37;  DB 11;  Length 293;
  Best Local Similarity   85.7%;  Pred. No. 17;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSH 7
              |||||:|
Db         86 GATCLTH 92


RESULT 12
Q7Z4Z0
ID   Q7Z4Z0      PRELIMINARY;      PRT;   310 AA.
AC   Q7Z4Z0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   N-terminal asparagine amidohydrolase.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jiang C.L., Yu L., Fan Y.X., Tu Q., Jiang J.X., Zhao S.Y.;
RT   "Cloning and expression of a novel human cDNA homology to murine N-
RT   terminal asparagine amidohydrolase (Ntan1) mRNA.";
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Zhao Y., Yu L., Xin Y.R., Zhang M., Chen S.Y., Zhao S.Y.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF092440; AAP97215.1; -.
KW   Hydrolase.
SQ   SEQUENCE   310 AA;  34691 MW;  44C92804CB4CA625 CRC64;

  Query Match             77.1%;  Score 37;  DB 4;  Length 310;
  Best Local Similarity   85.7%;  Pred. No. 18;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCLSH 7
              |||||:|
Db         86 GATCLTH 92


RESULT 13
O44143
ID   O44143      PRELIMINARY;      PRT;   669 AA.
AC   O44143;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   C44B12.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Tin-Wollam A.;
RT   "The sequence of C. elegans cosmid C44B12.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF036692; AAB88327.1; -.
DR   PIR; T32512; T32512.
DR   WormPep; C44B12.4; CE16920.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR   GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   669 AA;  76412 MW;  863406A1ED776AF0 CRC64;

  Query Match             77.1%;  Score 37;  DB 5;  Length 669;
  Best Local Similarity   66.7%;  Pred. No. 37;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GATCLSHSV 9
              |  |||||:
Db         62 GTLCLSHSI 70


RESULT 14
Q9GL46
ID   Q9GL46      PRELIMINARY;      PRT;   135 AA.
AC   Q9GL46;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Pref-1 (Fragment).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fahrenkrug S.C.;
RT   "Mapping of the porcine Pref-1 gene.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY007208; AAG09619.1; -.
DR   HSSP; P00750; 1TPG.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
KW   EGF-like domain.
FT   NON_TER       1      1
FT   NON_TER     135    135
SQ   SEQUENCE   135 AA;  14636 MW;  1AF406B033B24EA7 CRC64;

  Query Match             75.0%;  Score 36;  DB 6;  Length 135;
  Best Local Similarity   75.0%;  Pred. No. 13;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GATCLSHS 8
```

```
                | ||| ||
Db         31 GGTCLQHS 38


RESULT 15
O46370
ID   O46370      PRELIMINARY;      PRT;   308 AA.
AC   O46370;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Preadipocyte factor-1.
GN   DLK.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fat;
RA   Minoshima Y., Taniguchi Y., Sasaki Y.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Fahrenkrug S.C., Freking B.A., Smith T.P.S.;
RT   "Genomic Organization and Chromosomal Position of the Bovine DLK
RT   Gene.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB009278; BAA23733.1; -.
DR   EMBL; AF181466; AAF00926.1; -.
DR   EMBL; AF181463; AAF00926.1; JOINED.
DR   EMBL; AF181464; AAF00926.1; JOINED.
DR   EMBL; AF181465; AAF00926.1; JOINED.
DR   EMBL; AF181462; AAD56944.1; -.
DR   PIR; JC7125; JC7125.
DR   HSSP; P00740; 1EDM.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 4.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 5.
KW   EGF-like domain.
SQ   SEQUENCE   308 AA;  33005 MW;  FF42B9103DD9D852 CRC64;

  Query Match             75.0%;  Score 36;  DB 6;  Length 308;
  Best Local Similarity   75.0%;  Pred. No. 28;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GATCLSHS 8
              | ||| ||
Db        220 GGTCLQHS 227


Search completed: August 25, 2004, 02:31:21
Job time : 5.43416 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 25, 2004, 01:54:26 ; Search time 171.214 Seconds
(without alignments)
912.596 Million cell updates/sec

Title: US-09-593-793A-113
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2861	100.0	553	2	AAW69385	Aaw69385 Prostate
2	2861	100.0	553	2	AAW71869	Aaw71869 Amino aci
3	2861	100.0	553	3	AAY82002	Aay82002 Human imm
4	2861	100.0	553	3	AAB28527	Aab28527 Protein e
5	2861	100.0	553	3	ABG94411	Abg94411 Human pro
6	2861	100.0	553	4	AAM01117	Aam01117 Human pro
7	2861	100.0	553	4	AAU69763	Aau69763 Human pro
8	2861	100.0	553	4	AAB74800	Aab74800 Prostate
9	2861	100.0	553	4	AAG99002	Aag99002 Human pro
10	2861	100.0	553	4	AAG62150	Aag62150 Human P50
11	2861	100.0	553	4	ABU71653	Abu71653 Prostate
12	2861	100.0	553	4	AAU04961	Aau04961 Human pro
13	2861	100.0	553	5	AAU10324	Aau10324 Human PRO
14	2861	100.0	553	5	ABB95222	Abb95222 Human L1-
15	2861	100.0	553	5	AAU82643	Aau82643 Human bre
16	2861	100.0	553	5	ABG61900	Abg61900 Prostate
17	2861	100.0	553	5	ABG76665	Abg76665 Prostate
18	2861	100.0	553	5	ABB77575	Abb77575 Human mas
19	2861	100.0	553	6	ABR54334	Abr54334 Prostate
20	2861	100.0	553	7	ADB13563	Adb13563 Human pro
21	2861	100.0	1079	4	AAB74830	Aab74830 Prostate
22	2861	100.0	1079	4	ABU71860	Abu71860 Prostate
23	2601	90.9	710	5	AAM50661	Aam50661 Thioredox
24	2596	90.7	595	4	AAM01318	Aam01318 Alpha pre
25	1696	59.3	359	4	ABU71887	Abu71887 Human pro
26	1677.5	58.6	530	5	AAM50662	Aam50662 Thioredox
27	1517	53.0	305	6	ABR54583	Abr54583 Prostate
28	1517	53.0	305	7	ADB14479	Adb14479 Human P50
29	1417.5	49.5	371	4	AAM01230	Aam01230 P553S spl
30	1417.5	49.5	371	4	AAU69875	Aau69875 Human pro
31	1417.5	49.5	371	4	ABU71766	Abu71766 Prostate
32	1417.5	49.5	371	5	ABB95335	Abb95335 Human P55
33	1417.5	49.5	371	6	ABR54447	Abr54447 Prostate
34	1417.5	49.5	371	7	ADB14158	Adb14158 Human pro
35	1416	49.5	371	4	AAE01362	Aae01362 Human gen
36	1416	49.5	371	5	ABG64105	Abg64105 Human alb
37	1416	49.5	371	6	ADA57289	Ada57289 Human sec
38	1403.5	49.1	400	4	AAM01262	Aam01262 Ra12-P501
39	1403.5	49.1	400	4	AAU69907	Aau69907 Human pro
40	1403.5	49.1	400	4	ABU71798	Abu71798 Prostate
41	1403.5	49.1	400	5	ABB95367	Abb95367 Ra12-P501
42	1403.5	49.1	400	6	ABR54479	Abr54479 Prostate
43	1403.5	49.1	400	7	ADB14302	Adb14302 Human pro
44	1287	45.0	255	2	AAW85068	Aaw85068 Protein e
45	1287	45.0	255	3	AAB29268	Aab29268 Human pro

ALIGNMENTS

RESULT 1
AAW69385
ID AAW69385 standard; protein; 553 AA.
XX
AC AAW69385;
XX
DT 25-MAR-2003 (revised)
DT 08-DEC-1998 (first entry)
XX
DE Prostate tumour specific gene clone L1-12 protein.
XX
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy.
XX
OS Homo sapiens.
XX
PN WO9837418-A2.
XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US003690.
XX
PR 25-FEB-1997; 97US-00806596.
PR 01-AUG-1997; 97US-00904809.
PR 09-FEB-1998; 98US-00020747.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC;
XX
DR WPI; 1998-480805/41.
DR N-PSDB; AAV58586.
XX
PT Novel human prostate specific tumour protein and fragments - useful for
PT detecting and treating prostate cancers.
XX
PS Example 1; Page 87-89; 141pp; English.
XX
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as this
CC protein sequence). An antibody which binds to an immunogenic portion of
CC the prostate protein, and the method can be used to detect, monitor
CC progression of, or treat prostate cancers. The antibody may also be
CC conjugated to a therapeutic agent for use in therapy of prostate cancers.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX

SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 2; Length 553;
Best Local Similarity 100.0%; Pred. No. 2e-271;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 2
AAW71869
ID AAW71869 standard; protein; 553 AA.
XX
AC AAW71869;
XX
DT 06-JAN-1999 (first entry)
XX
DE Amino acid encoded by prostate tumour clone L1-12.
XX
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
XX
OS Homo sapiens.
XX
PN WO9837093-A2.
XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US003492.
XX
PR 25-FEB-1997; 97US-00806099.
PR 01-AUG-1997; 97US-00904804.
PR 09-FEB-1998; 98US-00020956.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC;
XX
DR WPI; 1998-609886/51.
DR N-PSDB; AAV61201.
XX
PT Polypeptides comprising immunogenic portions of prostate proteins - used
PT in a vaccine for the treatment of prostate cancer.
XX
PS Example 1; Page 82-84; 130pp; English.
XX
CC The present sequence is an immunogenic portion of a prostate tumour
CC protein. The immunogen, or the DNA encoding it, can be used as a vaccine
CC for the treatment of prostate cancer. The immunogen was isolated from a
CC prostate tumour cDNA library obtained by subtracting a prostate tumour
CC cDNA expression library with a normal tissue cDNA library
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 2; Length 553;
Best Local Similarity 100.0%; Pred. No. 2e-271;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 3
AAY82002
ID AAY82002 standard; protein; 553 AA.
XX
AC AAY82002;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein sequence SEQ ID NO:113.
XX

```
KW   Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW   immunogenic; cytostatic; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200004149-A2.
XX
PD   27-JAN-2000.
XX
PF   14-JUL-1999;   99WO-US015838.
XX
PR   14-JUL-1998;   98US-00115453.
PR   14-JUL-1998;   98US-00116134.
PR   23-SEP-1998;   98US-00159812.
PR   23-SEP-1998;   98US-00159822.
PR   15-JAN-1999;   99US-00232149.
PR   15-JAN-1999;   99US-00232880.
PR   09-APR-1999;   99US-00288946.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillon DC,  Harlocker SL,  Yuqiu J,  Xu J,  Mitcham JL;
XX
DR   WPI; 2000-171268/15.
XX
PT   New polypeptide useful for treating and diagnosing prostate cancer
PT   comprises an immunogenic portion of prostate tumor protein.
XX
PS   Claim 3; Page 138-139; 263pp; English.
XX
CC   The present invention describes isolated polypeptides, comprising an
CC   immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC   and polynucleotides encoding them have cytostatic activity and can be
CC   used in vaccines and in gene therapy. The polypeptides and
CC   polynucleotides encoding them, antigen presenting cells which express the
CC   polypeptides, antibodies against the polypeptides and vaccines comprising
CC   them can be used for inhibiting the development of prostate cancer in a
CC   patient. The polypeptides can be used to generate antibodies or anti-
CC   idiotypic antibodies for passive immuno therapy. A portion of the
CC   polynucleotides encoding the polypeptides can be used as a probe or to
CC   modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC   AAY82000 to AAY82020 represent sequences used in the exemplification of
CC   the present invention
XX
SQ   Sequence 553 AA;

  Query Match            100.0%;  Score 2861;  DB 3;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 2e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 4
AAB28527
ID   AAB28527 standard; protein; 553 AA.
XX
AC   AAB28527;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Protein encoded by human breast tumour cDNA clone P501S.
XX
KW   Human; breast tumour antigen; cytostatic; immunotherapy; breast cancer;
KW   vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200061756-A2.
XX
PD   19-OCT-2000.
XX
PF   10-APR-2000; 2000WO-US009688.
XX
PR   09-APR-1999;   99US-00288950.
PR   02-JUL-1999;   99US-00346327.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Xu J,  Dillon DC;
XX
DR   WPI; 2000-638568/61.
DR   N-PSDB; AAC79473.
XX
PT   A novel isolated polypeptide comprising an immunogenic portion of a
PT   breast cancer protein useful in the detection and treatment of breast
PT   cancer.
XX
PS   Claim 2; Page 92-93; 95pp; English.
XX
CC   The present sequence is encoded by a cDNA sequence which was isolated
CC   from a breast tumour cDNA library. It is provided in a specification
CC   relating to compounds for immunotherapy and diagnosis of breast cancer.
CC   Breast tumour antigens and the polynucleotides that encode them may be
CC   used in the production of a pharmaceutical composition to be used in the
CC   treatment of breast cancer. Proliferated T cells and incubated antigen
CC   presenting cells are also required. The polypeptides and polynucleotides
CC   may also be used to produce a vaccine
XX
SQ   Sequence 553 AA;

  Query Match            100.0%;  Score 2861;  DB 3;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 2e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
```

```
Qy          61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy         121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy         181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy         241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy         301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy         361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy         421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy         481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy         541 VVFDKSDLAKYSA 553
               |||||||||||||
Db         541 VVFDKSDLAKYSA 553
```

```
RESULT 5
ABG94411
ID   ABG94411 standard; protein; 553 AA.
XX
AC   ABG94411;
XX
DT   27-NOV-2002  (first entry)
XX
DE   Human prostate tumour protein partial sequence #3.
XX
KW   Human; immunogenic; prostate protein; prostate tumour protein;
KW   prostate cancer; cytostatic; vaccine.
XX
OS   Homo sapiens.
XX
PN   US2002090372-A1.
XX
PD   11-JUL-2002.
XX
PF   14-JUL-1998;   98US-00115453.
XX
PR   25-FEB-1997;   97US-00806099.
PR   01-AUG-1997;   97US-00904804.
PR   09-FEB-1998;   98US-00020956.
PR   25-FEB-1998;   98US-00030607.
XX
PA   (XUJJ/) XU J.
PA   (DILL/) DILLON D C.
XX
PI   Xu J,  Dillon DC;
XX
DR   WPI; 2000-171268/15.
DR   N-PSDB; ABS21254.
XX
PT   New polypeptide useful for treating and diagnosing prostate cancer
PT   comprises an immunogenic portion of prostate tumor protein.
XX
PS   Example 1; Page 59-61; 101pp; English.
XX
CC   The present invention relates to a new polypeptide comprising an
CC   immunogenic portion of a prostate protein. The invention is useful for
CC   inhibiting the development of prostate cancer in a patient. The invention
CC   is also useful as markers for diagnosing prostate cancer and for
CC   monitoring diseases progression in patients. The present amino acid
CC   sequence represents a human prostate tumour protein
XX
SQ   Sequence 553 AA;

  Query Match             100.0%;  Score 2861;  DB 3;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 2e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy           1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy          61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy         121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy         181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy         241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy         301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy         361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy         421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy         481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy         541 VVFDKSDLAKYSA 553
               |||||||||||||
Db         541 VVFDKSDLAKYSA 553
```

```
RESULT 6
AAM01117
ID   AAM01117 standard; protein; 553 AA.
XX
AC   AAM01117;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Human prostate-specific amino acid sequence L1-12.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis.
XX
OS   Homo sapiens.
XX
```

PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US001574.
XX
PR 14-JAN-2000; 2000US-00483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for diagnosing,
PT monitoring and treating prostate cancer in a patient and for use in
PT vaccines.
XX
PS Claim 2; Page 267-268; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated T cells
CC prepared using (I) or (II) are used treat cancer in a patient. (I) and
CC the antibodies are also used in the detection of cancer in a patient. The
CC cancer that is diagnosed or treated is particularly prostate cancer. (I)
CC and (II) can be used in vaccines. The antibodies or (I) can be used for
CC monitoring the progression of cancer in a patient. (I) and (II) can also
CC be used to improve diagnostic and therapeutic methods for prostate
CC cancer. They can indicate the level of metastasis as well as the prostate
CC volume. AAH93357 to AAH93944 and AAM01115 to AAM01318 represent
CC polynucleotide and amino acid sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 2e-271;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 7
AAU69763
ID AAU69763 standard; protein; 553 AA.
XX
AC AAU69763;
XX
DT 30-JAN-2002 (first entry)
XX
DE Human prostate cDNA encoded protein #3.
XX
KW Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX
OS Homo sapiens.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US009919.
XX
PR 27-MAR-2000; 2000US-00536857.
PR 09-MAY-2000; 2000US-00568100.
PR 12-MAY-2000; 2000US-00570737.
PR 13-JUN-2000; 2000US-00593793.
PR 27-JUN-2000; 2000US-00605783.
PR 09-AUG-2000; 2000US-00636215.
PR 29-AUG-2000; 2000US-00651236.
PR 06-SEP-2000; 2000US-00657279.
PR 02-OCT-2000; 2000US-00679426.
PR 10-OCT-2000; 2000US-00685166.
PR 09-NOV-2000; 2000US-00709729.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2001-639232/73.
DR N-PSDB; AAS63557.
XX
PT New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer.
XX
PS Claim 2; Page 269-270; 579pp; English.
XX
CC The invention relates to isolated prostate-specific polynucleotides,
CC polypeptides, fusion proteins of the polypeptides, antibodies raised
CC against the polypeptides (or antigenic epitopes derived from them) and
CC antigen-presenting cells expressing the polypeptides. The antibodies are
CC useful for detecting the presence of cancer, especially prostate cancer.
CC The polypeptides, polynucleotides and the antigen-presenting cells are
CC useful for stimulating and/or expanding T cells specific for a tumour
CC protein, and for inhibiting the development of cancer especially prostate
CC cancer. Compositions comprising the polynucleotide and/or polypeptide are
CC useful for stimulating an immune response, and for treating cancer. The
CC oligonucleotide is useful for detecting cancer. The present sequence is a
CC prostate specific polypeptide of the invention
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 4; Length 553;

```
  Best Local Similarity   100.0%;  Pred. No. 2e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 8
AAB74800
ID   AAB74800 standard; protein; 553 AA.
XX
AC   AAB74800;
XX
DT   14-JUN-2001  (first entry)
XX
DE   Prostate tumour antigen predicted amino acid sequence for L1-12.
XX
KW   Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;
KW   prostate cancer; immunogenic; cytostatic; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200125272-A2.
XX
PD   12-APR-2001.
XX
PF   04-OCT-2000; 2000WO-US027464.
XX
PR   04-OCT-1999;   99US-0157455P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Skeiky YAW,  Reed SG,  Cheever MA;
XX
DR   WPI; 2001-245062/25.
DR   N-PSDB; AAH02530.
XX
PT   Prostate specific protein and its encoding polynucleotide, useful for the
PT   treatment and diagnosis of prostate cancer.
XX
PS   Claim 3; Page 157-158; 276pp; English.
XX
CC   The present invention describes an isolated polypeptide (I) comprising at
CC   least an immunogenic portion of a prostate tumour antigen protein or its
CC   variant. (I) have cytostatic activity and can be used in vaccine
CC   production. (I), prostate tumour antigen polynucleotides, an antigen
CC   presenting cell (APC e.g. a dendritic cell) that expresses (I), and a
CC   pharmaceutical composition containing (I) are useful for inhibiting the
CC   development of cancer in a patient. Antibodies specific for prostate
CC   specific proteins and oligonucleotides that hybridise to a polynucleotide
CC   that encodes a prostate specific protein are useful for detecting the
CC   presence or absence of a cancer or monitoring the progression the
CC   progression of a cancer, especially prostate cancer. AAH02422 to AAH2872,
CC   AAB74798 to AAB74821 and AAB74830 are sequences used in the
CC   exemplification of the present invention
XX
SQ   Sequence 553 AA;

  Query Match             100.0%;  Score 2861;  DB 4;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 2e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 9
AAG99002
ID   AAG99002 standard; protein; 553 AA.
XX
AC   AAG99002;
XX
```

DT 25-SEP-2001 (first entry)
XX
DE Human prostate-specific amino acid sequence L1-12/P501S.
XX
KW Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW prostate specific antigen; PSA.
XX
OS Homo sapiens.
XX
PN WO200134802-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US030904.
XX
PR 12-NOV-1999; 99US-00439313.
PR 18-NOV-1999; 99US-00443686.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Retter MW, Stolk JA, Day CH, Skeiky YAW, Wang A;
XX
DR WPI; 2001-308785/32.
XX
PT Isolated polypeptide comprising at least an immunogenic portion of a
PT prostate-specific protein, useful in the diagnosis and therapy of
PT prostate cancer.
XX
PS Claim 3; Page 167-168; 325pp; English.
XX
CC The present invention describes an isolated polypeptide (P1) comprising
CC at least an immunogenic portion of a prostate-specific protein, or its
CC variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC (N1) have cytostatic activity and can be used in vaccine production. The
CC polypeptides, nucleic acids and antibodies from the present invention are
CC useful in the diagnosis and therapy of prostate cancer. Prostate specific
CC genes P704P, P712P, P774P, P775P and B305D are located in a genomic
CC region on chromosome 22q11.2 known as the Cat Eye Syndrome region.
CC Prostate specific antigen (PSA) P501S was located on chromosome 1.
CC AAH84671 to AAH85143 and AAG99000 to AAG99077 represent polynucleotide
CC and polypeptide sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 2e-271;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 10
AAG62150
ID AAG62150 standard; protein; 553 AA.
XX
AC AAG62150;
XX
DT 06-JUL-2001 (first entry)
XX
DE Human P501S inventive antigen SEQ ID NO: 333.
XX
KW Human; mouse; immunotherapy; cancer; leukaemia; WT1; Wilm's tumour gene;
KW chromosome 11p13; zinc finger transcription factor.
XX
OS Homo sapiens.
XX
PN WO200125273-A2.
XX
PD 12-APR-2001.
XX
PF 04-OCT-2000; 2000WO-US027465.
XX
PR 04-OCT-1999; 99US-0157459P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Xu J, Cheever MA, Reed SG;
XX
DR WPI; 2001-328324/34.
XX
PT Polypeptide comprising part of the Wilms Tumor gene product sequence is
PT used in the diagnosis and treatment of malignant diseases e.g. leukemia
PT and cancer associated with WTI.
XX
PS Disclosure; Page 212-213; 228pp; English.
XX
CC The present invention describes compositions comprising peptides derived
CC from the Wilm's tumour protein WT1 and methods for their use in treating
CC malignant diseases. Peptides derived from both the murine and human WT1
CC proteins are provided. The human WT1 gene is found on chromosome 11p13,
CC and the protein was shown to be a zinc finger transcription factor. The
CC immunogenic peptides of the invention are particularly useful in the
CC diagnosis and treatment of cancer and leukaemia. The present sequence is
CC a polypeptide described in the exemplification of the invention
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 2e-271;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
```

```
Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 11
ABU71653
ID ABU71653 standard; protein; 553 AA.
XX
AC ABU71653;
XX
DT 10-JUN-2003 (first entry)
XX
DE Prostate cancer specific antigen P501S.
XX
KW Prostate cancer; vaccine; gene therapy; cytostatic; fusion protein;
KW immunogen; cancer; prostate specific antigen.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2002192763-A1.
XX
PD 19-DEC-2002.
XX
PF 29-JUN-2001; 2001US-00895793.
XX
PR 04-OCT-1999; 99US-0157455P.
PR 04-OCT-2000; 2000US-00679272.
PR 28-MAR-2001; 2001US-00822827.
XX
PA (XUJJ/) XU J.
PA (DILL/) DILLON D C.
PA (MITC/) MITCHAM J L.
PA (HARL/) HARLOCKER S L.
PA (JIAN/) JIANG Y.
PA (KALO/) KALOS M D.
PA (FANG/) FANGER G R.
PA (RETT/) RETTER M W.
PA (STOL/) STOLK J A.
PA (DAYC/) DAY C H.
PA (VEDV/) VEDVICK T S.
PA (CART/) CARTER D.
PA (LISX/) LI S X.
PA (WANG/) WANG A.
PA (SKEI/) SKEIKY Y A W.
PA (HEPL/) HEPLER W T.
PA (HEND/) HENDERSON R A.
PA (HURA/) HURAL J.
PA (MCNE/) MCNEILL P D.
PA (HOUG/) HOUGHTON R L.
PA (DBAS/) Y DE BASSOLS C V.
PA (FOYT/) FOY T M.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA, Hural J;
PI Mcneill PD, Houghton RL, Y De BassolsCV, Foy TM;
XX
DR WPI; 2001-245062/25.
XX
PT Prostate specific protein and its encoding polynucleotide, useful for the
PT treatment and diagnosis of prostate cancer.
XX
PS Example 1; Fig 9; 85pp; English.
XX
CC The invention describes a fusion protein comprising at least one amino
CC acid sequence of immunogenic portions of any of the 3 sequences not
CC defined in the specification, or sequences having at least 70 or 90 %
CC sequence identity to any one of the 35 sequences defined in the USPTO web
CC site, which is encoded by any of the 4 nucleotide sequences not defined
CC in the specification. The fusion protein, composition and methods are
CC useful for diagnosing, preventing and/or treating cancer, particularly
CC prostate cancer. The proteins are useful as markers to indicate the
CC presence or absence of cancer. This is the amino acid sequence of a
CC prostate cancer specific antigen. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20020192763
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 2e-271;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
```

```
Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 12
AAU04961
ID   AAU04961 standard; protein; 553 AA.
XX
AC   AAU04961;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human prostate tumour protein L1-12.
XX
KW   Human; prostate tumour protein; prostate cancer.
XX
OS   Homo sapiens.
XX
PN   US6262245-B1.
XX
PD   17-JUL-2001.
XX
PF   25-FEB-1998;   98US-00030607.
XX
PR   25-FEB-1997;   97US-00806099.
PR   01-AUG-1997;   97US-00904804.
PR   09-FEB-1998;   98US-00020956.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC;
XX
DR   WPI; 2001-440862/47.
DR   N-PSDB; AAS10108.
XX
PT   Novel polynucleotide encoding polypeptide comprising a portion of
PT   prostate tumor protein useful for inhibiting development of prostate
PT   cancer or for treating prostate cancer in a patient.
XX
PS   Example 1; Col 125-127; 105pp; English.
XX
CC   The sequence is a partial prostate tumour protein, encoded by a prostate
CC   tumour specific cDNA. The DNA is useful for inhibiting the development of
CC   prostate cancer or for treating prostate cancer in a patient
XX
SQ   Sequence 553 AA;

  Query Match           100.0%;  Score 2861;  DB 4;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 2e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 13
AAU10324
ID   AAU10324 standard; protein; 553 AA.
XX
AC   AAU10324;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Human PROST 03.
XX
KW   Human; PROST 03; metastasis; prostate cancer; tumour; immune response;
KW   cytostatic; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200181577-A2.
XX
PD   01-NOV-2001.
XX
PF   26-APR-2001; 2001WO-US013323.
XX
PR   27-APR-2000; 2000US-0200065P.
PR   20-APR-2001; 2001US-00838785.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Lau T,  Lin RJ,  Parkes D,  Parry G,  Schneider DW,  Steinbrecher R;
PI   Van Heuit PT,  Wu J;
XX
DR   WPI; 2002-041404/05.
DR   N-PSDB; AAS14962.
XX
PT   Novel PROST 03 polypeptides and polynucleotides useful in research,
PT   diagnosis and therapeutic applications, particularly for use in cancer
PT   therapeutics.
XX
PS   Claim 14; Fig 2; 77pp; English.
XX
CC   The invention relates to an isolated PROST 03 polypeptide (I) and to the
CC   polynucleotide (II) encoding PROST 03. Fragments of (I) were used to
CC   generate antibodies (III) to PROST 03. (III) is useful for selectively
CC   destroying a cell expressing (I), and for treating a disease-state
CC   associated with expression of PROST 03 in a human patient. (III) is
CC   useful for diagnosing metastasis associated with (I), in a subject. (I)
CC   is also useful for diagnosing and treating diseases of cell proliferation
```

```
CC   such as prostate cancers. (I) is also useful for generating antibodies to
CC   PROST 03. (III) is useful in detecting the levels of PROST 03
CC   polypeptides in cells and tissues, and in targeting drugs to primary and
CC   metastatic tumours. (I) is also useful for stimulating immune response to
CC   PROST 03 containing cells. (II) is useful in diagnostic assays for
CC   detecting the levels of polynucleotides encoding PROST 03 in cells and
CC   tissues. (II) is useful as DNA probes, as targets for antisense and
CC   ribozyme therapy, and as templates for the production of antisense
CC   polynucleotides. (I) and (II) are useful in research, biological,
CC   clinical and therapeutic purposes. The present sequence represents the
CC   amino acid sequence of human PROST 03
XX
SQ   Sequence 553 AA;

  Query Match           100.0%;  Score 2861;  DB 5;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 2e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 14
ABB95222
ID   ABB95222 standard; protein; 553 AA.
XX
AC   ABB95222;
XX
DT   19-JUL-2002  (first entry)
XX
DE   Human L1-12 protein SEQ ID NO 113.
XX
KW   Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
PN   US2002022248-A1.
XX
PD   21-FEB-2002.
XX
PF   12-JAN-2001; 2001US-00759143.
XX
PR   25-FEB-1997;   97US-00806099.
PR   01-AUG-1997;   97US-00904804.
PR   10-FEB-1998;   98US-00020956.
PR   25-FEB-1998;   98US-00030607.
PR   14-JUL-1998;   98US-00115453.
PR   23-SEP-1998;   98US-00159812.
PR   15-JAN-1999;   99US-00232149.
PR   09-APR-1999;   99US-00288946.
PR   13-JUL-1999;   99US-00352616.
PR   12-NOV-1999;   99US-00439313.
PR   18-NOV-1999;   99US-00443686.
PR   14-JAN-2000; 2000US-00483672.
PR   27-MAR-2000; 2000US-00536857.
PR   09-MAY-2000; 2000US-00568100.
PR   12-MAY-2000; 2000US-00570737.
PR   13-JUN-2000; 2000US-00593793.
PR   27-JUN-2000; 2000US-00605783.
PR   10-AUG-2000; 2000US-00636215.
PR   29-AUG-2000; 2000US-00651236.
PR   06-SEP-2000; 2000US-00657279.
PR   02-OCT-2000; 2000US-00679426.
PR   10-OCT-2000; 2000US-00685166.
XX
PA   (XUJJ/) XU J.
PA   (DILL/) DILLON D C.
PA   (MITC/) MITCHAM J L.
PA   (HARL/) HARLOCKER S L.
PA   (JIAN/) JIANG Y.
PA   (KALO/) KALOS M D.
PA   (FANG/) FANGER G R.
PA   (RETT/) RETTER M W.
PA   (STOL/) STOLK J A.
PA   (DAYC/) DAY C H.
PA   (VEDV/) VEDVICK T S.
PA   (CART/) CARTER D.
PA   (LISX/) LI S X.
PA   (WANG/) WANG A.
PA   (SKEI/) SKEIKY Y A W.
PA   (HEPL/) HEPLER W T.
PA   (HEND/) HENDERSON R A.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Kalos MD;
PI   Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS,  Carter D;
PI   Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Henderson RA;
XX
DR   WPI; 2002-255649/30.
XX
PT   New prostate-specific polynucleotides for diagnosing and treating
PT   diseases, in particular prostate cancer, and as markers for the
PT   progression of cancer.
XX
PS   Claim 2; SEQ ID NO 113; 87pp; English.
XX
CC   The present invention provides prostate-specific coding sequences and
CC   their encoded proteins. These can be used in the diagnosis and treatment
CC   of cancers, particularly prostate cancer. The present sequence is a
CC   protein described in the invention
XX
SQ   Sequence 553 AA;

  Query Match           100.0%;  Score 2861;  DB 5;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 2e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 15
AAU82643
ID AAU82643 standard; peptide; 553 AA.
XX
AC AAU82643;
XX
DT 23-APR-2002 (first entry)
XX
DE Human breast tumour polypeptide clone #3.
XX
KW Human; breast tumour polypeptide; breast cancer; cytostatic;
KW immunostimulant.
XX
OS Homo sapiens.
XX
PN WO200198339-A2.
XX
PD 27-DEC-2001.
XX
PF 12-JUN-2001; 2001WO-US019032.
XX
PR 22-JUN-2000; 2000US-00602877.
PR 12-OCT-2000; 2000US-00687507.
PR 06-FEB-2001; 2001US-00778381.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Xu J, Dillon DC, Retter MW, Harlocker SL;
XX
DR WPI; 2002-147792/19.
DR N-PSDB; ABK29012.
XX
PT Polynucleotides encoding breast tumor polypeptides, useful for treating
PT breast cancer or stimulating an immune response.
XX
PS Claim 2; Page 144-145; 150pp; English.
XX
CC The invention relates to polynucleotides encoding breast tumour
CC polypeptides. The sequences are useful for treating cancer, preferably
CC breast cancer, in a patient or for stimulating an immune response. The
CC polynucleotides and polypeptides are also useful in the diagnosis and
CC monitoring of breast cancer. A method for detecting the presence of a
CC cancer in a patient, comprises obtaining a biological sample from the
CC patient, contacting the biological sample with a binding agent that binds
CC to a breast tumour polypeptide, detecting in the sample an amount of
CC polypeptide that binds to the binding agent, and comparing the amount of
CC polypeptide to a predetermined cut-off value, therefore determining the
CC presence of a cancer in the patient. Sequences AAU82641-AAU82655
CC represent human breast tumour polypeptides of the invention
XX
SQ Sequence 553 AA;

```
  Query Match             100.0%;  Score 2861;  DB 5;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 2e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

Search completed: August 25, 2004, 02:27:08
Job time : 175.214 secs

This Page Blank (usp)

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 25, 2004, 02:24:18 ; Search time 162.358 Seconds
(without alignments)
1070.372 Million cell updates/sec

Title: US-09-593-793A-113
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1295152 seqs, 314255058 residues

Total number of hits satisfying chosen parameters: 1295152

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2861	100.0	553	9	US-09-745-288-101	Sequence 101, App
2	2861	100.0	553	9	US-09-838-785-2	Sequence 2, Appli
3	2861	100.0	553	9	US-09-759-143-113	Sequence 113, App
4	2861	100.0	553	9	US-09-780-669-113	Sequence 113, App
5	2861	100.0	553	9	US-09-030-606-113	Sequence 113, App
6	2861	100.0	553	9	US-09-822-827-113	Sequence 113, App
7	2861	100.0	553	9	US-09-115-453-113	Sequence 113, App
8	2861	100.0	553	9	US-09-232-880-113	Sequence 113, App
9	2861	100.0	553	9	US-09-895-793-113	Sequence 113, App
10	2861	100.0	553	9	US-09-895-814-113	Sequence 113, App
11	2861	100.0	553	12	US-10-453-919-101	Sequence 101, App
12	2861	100.0	553	13	US-10-012-896-113	Sequence 113, App
13	2861	100.0	553	14	US-10-010-940-113	Sequence 113, App
14	2861	100.0	553	14	US-10-144-678A-113	Sequence 113, App
15	2861	100.0	553	14	US-10-005-907-13	Sequence 13, Appl
16	2861	100.0	553	14	US-10-294-025-113	Sequence 113, App
17	2861	100.0	553	15	US-10-295-027-548	Sequence 548, App
18	2861	100.0	553	15	US-10-295-027-902	Sequence 902, App
19	2861	100.0	553	16	US-10-688-838-113	Sequence 113, App
20	2861	100.0	1079	9	US-09-822-827-947	Sequence 947, App
21	2861	100.0	1079	9	US-09-895-793-947	Sequence 947, App
22	2601	90.9	710	14	US-10-296-770-4	Sequence 4, Appli
23	1696	59.3	359	9	US-09-822-827-974	Sequence 974, App
24	1696	59.3	359	9	US-09-895-793-974	Sequence 974, App
25	1677.5	58.6	530	14	US-10-296-770-5	Sequence 5, Appli
26	1517	53.0	305	14	US-10-144-678A-1029	Sequence 1029, Ap
27	1517	53.0	305	14	US-10-294-025-1029	Sequence 1029, Ap
28	1417.5	49.5	371	9	US-09-759-143-708	Sequence 708, App
29	1417.5	49.5	371	9	US-09-780-669-708	Sequence 708, App
30	1417.5	49.5	371	9	US-09-822-827-708	Sequence 708, App
31	1417.5	49.5	371	9	US-09-895-793-708	Sequence 708, App
32	1417.5	49.5	371	9	US-09-895-814-708	Sequence 708, App
33	1417.5	49.5	371	13	US-10-012-896-708	Sequence 708, App
34	1417.5	49.5	371	14	US-10-144-678A-708	Sequence 708, App
35	1417.5	49.5	371	14	US-10-294-025-708	Sequence 708, App
36	1416	49.5	371	11	US-09-833-245-852	Sequence 852, App
37	1403.5	49.1	400	9	US-09-759-143-852	Sequence 852, App
38	1403.5	49.1	400	9	US-09-780-669-852	Sequence 852, App
39	1403.5	49.1	400	9	US-09-822-827-852	Sequence 852, App
40	1403.5	49.1	400	9	US-09-895-793-852	Sequence 852, App
41	1403.5	49.1	400	9	US-09-895-814-852	Sequence 852, App
42	1403.5	49.1	400	13	US-10-012-896-852	Sequence 852, App
43	1403.5	49.1	400	14	US-10-144-678A-852	Sequence 852, App
44	1403.5	49.1	400	14	US-10-294-025-852	Sequence 852, App
45	1287	45.0	255	12	US-09-841-894-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-745-288-101
; Sequence 101, Application US/09745288
; Patent No. US20010018058A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.446D1
; CURRENT APPLICATION NUMBER: US/09/745,288
; CURRENT FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-745-288-101

Query Match 100.0%; Score 2861; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
```

```
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553

Db        541 VVFDKSDLAKYSA 553
```

RESULT 2
US-09-838-785-2
; Sequence 2, Application US/09838785
; Patent No. US20020009455A1
; GENERAL INFORMATION:
; APPLICANT: Lau, Ted
; APPLICANT: Lin, Rick
; APPLICANT: Parkes, Debbie
; APPLICANT: Parry, Gordon
; APPLICANT: Schneider, Douglas
; APPLICANT: Steinbrecher, Renate
; APPLICANT: Van Heuit, Pam T
; APPLICANT: Wu, John
; TITLE OF INVENTION: DNA Encoding a No. US20020009455A1el PROST 03
; FILE REFERENCE: 51831AUSM1
; CURRENT APPLICATION NUMBER: US/09/838,785
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/200,065
; PRIOR FILING DATE: 2000-04-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-838-785-2

Query Match 100.0%; Score 2861; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553

Db        541 VVFDKSDLAKYSA 553
```

RESULT 3
US-09-759-143-113
; Sequence 113, Application US/09759143
; Patent No. US20020022248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-759-143-113

Query Match 100.0%; Score 2861; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
```

```
Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
Db        541 VVFDKSDLAKYSA 553
```

RESULT 4
US-09-780-669-113
; Sequence 113, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780,669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-780-669-113

Query Match 100.0%; Score 2861; DB 9; Length 553;

Best Local Similarity 100.0%; Pred. No. 3.7e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
Db        541 VVFDKSDLAKYSA 553
```

RESULT 5
US-09-030-606-113
; Sequence 113, Application US/09030606
; Patent No. US20020081580A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF PROSTATE CANCER AND METHODS I
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,606
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.428C3

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  113:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-606-113

  Query Match             100.0%;  Score 2861;  DB 9;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-245;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 6
US-09-822-827-113
; Sequence 113, Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.534C1
;  CURRENT APPLICATION NUMBER: US/09/822,827
;  CURRENT FILING DATE:  2001-03-28
;  NUMBER OF SEQ ID NOS: 982
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-822-827-113

  Query Match             100.0%;  Score 2861;  DB 9;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-245;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 7
US-09-115-453-113
; Sequence 113, Application US/09115453B
; Patent No. US20020090372A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT: Dillon, Davin C.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND
;  TITLE OF INVENTION: METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.427C4
;  CURRENT APPLICATION NUMBER: US/09/115,453B
;  CURRENT FILING DATE: 1998-07-14
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-115-453-113

  Query Match             100.0%;  Score 2861;  DB 9;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-245;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 8
US-09-232-880-113
; Sequence 113, Application US/09232880
; Publication No. US20020182596A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF
; TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.428C6
; CURRENT APPLICATION NUMBER: US/09/232,880
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-232-880-113

Query Match 100.0%; Score 2861; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 9
US-09-895-793-113
; Sequence 113, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien

US-09-895-793-113

Query Match 100.0%; Score 2861; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 10
US-09-895-814-113
; Sequence 113, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-895-814-113

Query Match 100.0%; Score 2861; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 11
US-10-453-919-101
; Sequence 101, Application US/10453919
; Publication No. US20040033230A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.446C7
; CURRENT APPLICATION NUMBER: US/10/453,919

```
;  CURRENT FILING DATE:  2003-06-03
;  NUMBER OF SEQ ID NOS: 121
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 101
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-453-919-101

  Query Match            100.0%;  Score 2861;  DB 12;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-245;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 12
US-10-012-896-113
; Sequence 113, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Hepler, William T.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Hural, John
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  Vinals de Bassols, Carlota
;  APPLICANT:  Foy, Teresa
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Wantanabe, Yoshihiro
;  APPLICANT:  Meagher, Madeleine Joy
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C27
;  CURRENT APPLICATION NUMBER: US/10/012,896
;  CURRENT FILING DATE:  2001-12-10
;  NUMBER OF SEQ ID NOS: 1011
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-012-896-113

  Query Match            100.0%;  Score 2861;  DB 13;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-245;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 13
US-10-010-940-113
; Sequence 113, Application US/10010940
; Publication No. US20030088062A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427D3
;  CURRENT APPLICATION NUMBER: US/10/010,940
;  CURRENT FILING DATE:  2001-12-05
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-010-940-113

  Query Match           100.0%;  Score 2861;  DB 14;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-245;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 14
US-10-144-678A-113
; Sequence 113, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A. W.
;  APPLICANT:  Hepler, William T.
;  APPLICANT:  Hural, John
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Houghton, Raymond L.
;  APPLICANT:  Vinals y de Bassols, Carlota
;  APPLICANT:  Foy, Teresa M.
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Deng, Ta
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C28
;  CURRENT APPLICATION NUMBER: US/10/144,678A
;  CURRENT FILING DATE:  2002-08-12
;  NUMBER OF SEQ ID NOS: 1033
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-144-678A-113

  Query Match           100.0%;  Score 2861;  DB 14;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-245;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
```

```
Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

RESULT 15
US-10-005-907-13
; Sequence 13, Application US/10005907
; Publication No. US20030166881A1
; GENERAL INFORMATION:
; APPLICANT: Union Chimique Belge, S.A.
; APPLICANT: No. US20030166881A1ka, Karl
; APPLICANT: Pirozzi, Gregory
; APPLICANT: Einstein, Richard
; TITLE OF INVENTION: NOVEL GENES ASSOCIATED WITH ALLERGIC HYPERSENSITIVITY AND MAST CE
; TITLE OF INVENTION: ACTIVATION
; FILE REFERENCE: 053529-5005
; CURRENT APPLICATION NUMBER: US/10/005,907
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-005-907-13

Query Match 100.0%; Score 2861; DB 14; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVQRLWVSRLLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553
```

Search completed: August 25, 2004, 02:34:07
Job time : 164.358 secs

This page blank (usp

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 25, 2004, 02:00:53 ; Search time 49.1993 Seconds
(without alignments)
1081.193 Million cell updates/sec

Title: US-09-593-793A-113
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	347.5	12.1	515	2	T14340	sucrose-proton tra
2	337.5	11.8	523	2	T12198	sucrose transport
3	330	11.5	516	2	JQ2389	sucrose transport
4	328.5	11.5	594	2	G84441	probable sucrose-p
5	324.5	11.3	525	2	S28052	sucrose transport
6	323.5	11.3	512	2	S38196	sucrose transport
7	323.5	11.3	512	2	G86360	probable sucrose-p
8	317.5	11.1	507	2	S48789	sucrose transport
9	311	10.9	533	2	S43142	sucrose transport
10	305.5	10.7	501	2	T14339	sucrose-proton tra
11	297.5	10.4	428	2	S48788	sucrose transport
12	294	10.3	510	2	S38657	sucrose transport
13	294	10.3	523	2	S51114	sucrose-proton sym
14	292.5	10.2	512	2	F96741	probable sucrose t
15	292	10.2	537	2	T02982	probable sucrose t
16	286	10.0	474	2	A86234	hypothetical prote
17	281.5	9.8	491	2	E96691	probable sucrose-p
18	279.5	9.8	513	2	S38197	sucrose transport
19	276.5	9.7	492	2	A84520	probable sucrose-p
20	273	9.5	503	2	S52377	sucrose transport
21	235.5	8.2	553	2	T38541	probable sucrose c
22	191	6.7	452	2	F75217	hypothetical prote
23	187.5	6.6	541	2	B87532	transporter, proba
24	142	5.0	544	2	S75696	melibiose carrier
25	139.5	4.9	454	2	A75444	hypothetical prote
26	138.5	4.8	430	2	E75217	transporter PAB217
27	136	4.8	418	2	B87536	membrane protein,
28	133	4.6	389	2	G83413	probable MFS trans
29	128.5	4.5	422	2	G83503	probable MFS trans
30	128.5	4.5	594	2	A83096	probable permease
31	127.5	4.5	472	2	F82639	resistance protein
32	127	4.4	451	2	E81781	probable integral
33	126	4.4	394	2	AB3073	tetracycline resis
34	126	4.4	394	2	G98213	drug efflux protei
35	126	4.4	451	2	A81206	sugar transporter,
36	125.5	4.4	399	2	JQ1479	tetracycline resis
37	125	4.4	391	2	S74688	hypothetical prote
38	125	4.4	407	2	H75515	tetracycline-efflu
39	124.5	4.4	399	1	S38656	tetA protein - Pse
40	123.5	4.3	400	2	AG2866	MFS permease [drug
41	123.5	4.3	400	2	D97643	probable efflux pr
42	123	4.3	440	2	AD0986	hypothetical metab
43	121.5	4.2	398	2	C90349	multidrug-efflux t
44	121.5	4.2	503	2	S43017	puromycin resistan
45	121	4.2	483	2	AF2363	hypothetical prote

ALIGNMENTS

RESULT 1
T14340
sucrose-proton transport protein - carrot
N;Alternate names: sucrose/H+ symporter protein
C;Species: Daucus carota (carrot)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T14340
R;Shakya, R.; Sturm, A.
Plant Physiol. 118, 1473-1480, 1998
A;Title: Characterization of source- and sink-specific sucrose/H+ symporters from carro
A;Reference number: Z17991; MUID:99063785; PMID:9847123
A;Accession: T14340
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-515 <SHA>
A;Cross-references: EMBL:Y16768; NID:g2969883; PIDN:CAA76369.1; PID:g2969884
A;Experimental source: cultivar Namtaise; root
C;Genetics:
A;Note: SUT2
C;Superfamily: common tobacco sucrose transport protein

Query Match 12.1%; Score 347.5; DB 2; Length 515;
Best Local Similarity 25.9%; Pred. No. 1e-18;
Matches 131; Conservative 80; Mismatches 215; Indels 79; Gaps 17;

```
Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
              :|:||  :  |::  |  :: : | :  :|:  |:  :   ||: |::  |::|  ||
Db         34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
              | :  :|||||||| : :  : :|: ||   |: | :      :    :|  : : ::|
Db         94 HCQSSFGRRRPFIASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIGF 150

Qy        130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
              :||    :   |  |||:||   |    | | : |:| :::|  |||          |
Db        151 WILDVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLF 210

Qy        179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
              |    | |   |    : |    : |: :  | |  :| | :  || : |:    |
Db        211 PFS--KTHACDLYCANLKSCFIISIALLIIITVVALSVVRENS--GPPDDADAAEEP--- 263

Qy        239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
              |   ::     | ||  | |    || :  | :    :|:|   | || ||::|
Db        264 ----PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGR 313

Qy        299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
              :| |   | |    : ||:||| |:||| |   :  : |: :: ||:  |   : |
Db        314 EIYGGT--AGQG----KLYDQGVRAGALGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWG 366

Qy        359 VAAFPVAAGATCL----------SHSV--------AVVTASA----ALTGFTFSALQIL 395
              | :| |                ||          | | | |    ::  |   |    :
Db        367 FVNFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSI 426
```

```
Qy        396 PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGA 448
Db        427 PFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISA 486

Qy        449 GGSGLLP----PPPALCGASACDVS 469
Db        487 AISGVLAIVLLPKPSKDAASKLSLS 511


RESULT 2
T12198
sucrose transport protein - fava bean
C;Species: Vicia faba (fava bean)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 20-Jun-2000
C;Accession: T12198
R;Weber, H.; Borisjuk, L.; Heim, U.; Sauer, N.; Wobus, U.
Plant Cell 9, 895-908, 1997
A;Title: A role for sugar transporters during seed development: molecular characterizati
A;Reference number: Z17451; MUID:97355984; PMID:9212465
A;Accession: T12198
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-523 <WEB>
A;Cross-references: EMBL:Z93774
C;Genetics:
A;Gene: sut
C;Superfamily: common tobacco sucrose transport protein

  Query Match             11.8%;  Score 337.5;  DB 2;  Length 523;
  Best Local Similarity   25.8%;  Pred. No. 5.8e-18;
  Matches  134;  Conservative   76;  Mismatches  185;  Indels  125;  Gaps   20;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
Db         36 KIMVVASIAAGVQFGWALQLSLLTPYVQLLGIHHTWAAYIWLCGPISGMLVQPIVGYHSD 95

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
Db         96 RCTSRFGRRRPFIAAGSIAVAIAVFLI---GYAADLGHSFGDSLDQKVRPRAIGIFVVGF 152

Qy        130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID------ 182
Db        153 WILDVANNMLQGPCRALLGDLCAGNQRKTRNANAFFSFFMAVGNVLGYAAGAYSKLYHVF 212

Qy        183 --WDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLV-AEEAALGPTEPA----EGLSAP 235
Db        213 PFTKTKACNVYCANLKSCFF--LSIALLTVLATSALIYVKETALTPEKTVVTTEDGGSSG 270

Qy        236 SLSPHCCPCRARL--AFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293
Db        271 GM-----PCFGQLSGAFKEL-------------KRPMWILLLVTCLNWIAWFPFLLFDT 311

Qy        294 DFVGEGLYQGVPRAEPGTEARRH-YDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTR 352
Db        312 DWMGKEVY-------GGTVGEGHAYDMGVREGALGLMLNSVVLGATSLGVDILARGVG-G 363

Qy        353 AVYLASVAAFPVAAGATCLSHSVAVV-------------------------TASAALTGF 387
Db        364 VKRLWGIVNFLL---AICLGLTVLVTKLAQHSRQYAPGTGALGDPLPPSEGIKAGALTLF 420

Qy        388 TFSALQI-----LPYTLASLYHREKQVFLPKYRGDTGGASSEDSL------------MTS 430
Db        421 SVLGVPLAITYSIPFALASIF------------SSTSGAGQGLSLGVLNLAIVIPQMFVS 468

Qy        431 FLPGPKPG-------APFPNGHVGAGGSG-----LLPPPP 458
Db        469 VLSGPWDALFGGGNLPAFVVGAVAALASGILSIILLPSPP 508


RESULT 3
JQ2389
sucrose transport protein - potato
C;Species: Solanum tuberosum (potato)
C;Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 21-Jul-2000
C;Accession: JQ2389; S40310
R;Riesmeier, J.W.; Hirner, B.; Frommer, W.B.
Plant Cell 5, 1591-1598, 1993
A;Title: Potato sucrose transporter expression in minor veins indicates a role in phloem
A;Reference number: JQ2389; MUID:94146554; PMID:8312741
A;Accession: JQ2389
A;Molecule type: mRNA
A;Residues: 1-516 <RIE>
A;Cross-references: EMBL:X69165; NID:g439293; PIDN:CAA48915.1; PID:g439294
A;Experimental source: cv. Desiree
C;Comment: The gene encoding for this protein is highly expressed in mature leaves.
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: glycoprotein; transmembrane protein
F;31-53/Domain: transmembrane #status predicted <TM1>
F;67-86/Domain: transmembrane #status predicted <TM2>
F;103-122/Domain: transmembrane #status predicted <TM3>
F;141-160/Domain: transmembrane #status predicted <TM4>
F;180-200/Domain: transmembrane #status predicted <TM5>
F;226-248/Domain: transmembrane #status predicted <TM6>
F;285-304/Domain: transmembrane #status predicted <TM7>
F;331-349/Domain: transmembrane #status predicted <TM8>
F;366-385/Domain: transmembrane #status predicted <TM9>
F;409-427/Domain: transmembrane #status predicted <TM10>
F;429-448/Domain: transmembrane #status predicted <TM11>
F;3,92/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             11.5%;  Score 330;  DB 2;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 2.1e-17;
  Matches  132;  Conservative   93;  Mismatches  197;  Indels  112;  Gaps   21;

Qy          4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
Db         29 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 78

Qy         64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPD 116
Db         79 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVMIAVFLI---GFAADLGHASGDTLGKG 135

Qy        117 PRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLG 175
Db        136 FKPRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGRMRTANAFFSFFMAVGNILG 195

Qy        176 YLLPAIDW--------DTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEE-AALGP 225
Db        196 YAAGSYSHLFKVFPFSKTKACDMYCANLKSCFFIAIFLLLSLTTIALTLVRENELPEKDE 255

Qy        226 TEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMAL 285
Db        256 QEIDEKLAGAG--------KSKVPF--FGEIFGALKEL----PRPMWILLLVTCLNWIAW 301

Qy        286 MTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRL 345
Db        302 FPFFLYDTDWMAKEVFGG----QVGD--ARLYDLGVRAGAMGLLLQSVVLGFMSLGVEFL 355

Qy        346 VQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT-------- 388
Db        356 GKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRQHDPAGTLMGPTPGVKIGAL 411

Qy        389 --FSALQI-------LPYTLASLYHREK------------QVFLPKYRGDTGGASSEDS 426
Db        412 LLFAALGIPLAATFSIPFALASIFSSNRGSGQGLSLGVLNLAIVVPQMLVSLVGGPWDDL 471

Qy        427 LMTSFLPGPKPGAPFPNGHVGAGGSG-----LLPPPPALCGASACDVSVRVVVG 475
Db        472 FGGGNLPG------FVVGAVAAAASAVLALTMLPSPPA-------DAKPAVAMG 512


RESULT 4
G84441
probable sucrose-proton symporter [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
```

```
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: G84441
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: G84441
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-594 <STO>
A;Cross-references: GB:AE002093; NID:g3461813; PIDN:AAC32907.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g02860
A;Map position: 2
C;Superfamily: common tobacco sucrose transport protein

  Query Match             11.5%;  Score 328.5;  DB 2;  Length 594;
  Best Local Similarity   24.1%;  Pred. No. 3.2e-17;
  Matches  120;  Conservative   72;  Mismatches  174;  Indels  131;  Gaps   18;

Qy          22 NLLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
Db          59 SLVTLVLSCTVAAGVQFGWALQLSLLTPYIQTLGISHAFSSFIWLCGPITGLVVQPFVGI 118

Qy          74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPD----------PRPLE 121
Db         119 WSDKCTSKYGRRRPFILVGSFMISIAVIII---GFSADIGYLLGDSKEHCSTFKGTRTRA 175

Qy         122 LALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAI 181
Db         176 AVVFIIGFWLLDLANNTVQGPARALLADL-SGPDQRNTANAVFCLWMAIGNILGFSAGAS 234

Qy         182 ----DWDTSALAPYLGTQEECLFG--------LLTLIFLT-CVAATLLVAEEAALGPTEP 228
Db         235 GKWQEW-----FPFL-TSRACCAACGNLKAAFLLAVVFLTICTLVTIYFAKEIPFTSNKP 288

Qy         229 AE-GLSAPSLSPHCCPCRARLAFRNL---------------------------------- 253
Db         289 TRIQDSAPLLDD--------LQSKGLEHSKLNNGTANGIKYERVERDTDEQFGNSENEHQ 340

Qy         254 --------GALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVP 305
Db         341 DETYVDGPGSVLVNLLTSLRHLPPAMHSVLIVMALTWLSWFPFFLFDTDWMGREVYHGDP 400

Qy         306 RAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVY-LASVAAFPV 364
Db         401 TGD--SLHMELYDQGVREGALGLLLNSVVLGISSFLIEPMCQRMGARVVWALSNFTVFAC 458

Qy         365 AAGATCLS--------------------HSVAVVTASAALTGFTFSALQILPYTLASLY 403
Db         459 MAGTAVISLMSLSDDKNGIEYIMRGNETTRTAAVIV--FALLGFPLAITYSVPFSVTA-- 514

Qy         404 HREKQVFLPKYRGDTGG 420
Db         515 ---------EVTADSGG 522


RESULT 5
S28052
sucrose transport protein - spinach
N;Alternate names: sucrose carrier protein; sucrose permease
C;Species: Spinacia oleracea (spinach)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000
C;Accession: S28052
R;Riesmeier, J.W.; Willmitzer, L.; Frommer, W.B.
EMBO J. 11, 4705-4713, 1992
A;Title: Isolation and characterization of a sucrose carrier cDNA from spinach by functi
A;Reference number: S28052; MUID:93099843; PMID:1464305
A;Accession: S28052
A;Molecule type: mRNA
A;Residues: 1-525 <RIE>
A;Cross-references: EMBL:X67125; NID:g21318; PIDN:CAA47604.1; PID:g21319
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: transmembrane protein

  Query Match             11.3%;  Score 324.5;  DB 2;  Length 525;
  Best Local Similarity   23.2%;  Pred. No. 5.7e-17;
  Matches  129;  Conservative   91;  Mismatches  196;  Indels  141;  Gaps   20;

Qy          23 LLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74
Db          35 LKKLGLVASVAAGVQFGWALQLSLLTPYVQLLGIPHTWAAYIWLCGPISGMIVQPLVGYY 94

Qy          75 SDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPDP-----RPLELALLIL 127
Db          95 SDRCTSRFGRRRPFIAAGAALVAVAVGLI---GFAADIGAASGDPTGNVAKPRAIAVFVV 151

Qy         128 GVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID---- 182
Db         152 GFWILDVANNTLQGPCRALLADMAAGSQTKTRYANAFFSFFMALGNIGGYAAGSYSRLYT 211

Qy         183 ----WDTSALAPYLGTQEECLFGLLT-LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
Db         212 VFPFTKTAACDVYCANLKSCFFISITLLIVLTILALSVVKERQITIDEIQEEEDLKNRNN 271

Qy         238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
Db         272 SSGC----ARLPF--FGQLIGALKDL----PKPMLILLLVTALNWIAWFPFLLFDTDWMG 321

Qy         298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVY 355
Db         322 KEVYGGT------VGEGKLYDQGVHAGALGLMINSVVLGVMSLSIEGLARMVGGAKRLWG 375

Qy         356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415
Db         376 IVNIIL------AVCLAMTV-LVTKSAE------------------HFRDSHHIM---- 405

Qy         416 GDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVG 475
Db         406 ---------------------------------GSAVPPPPPA--GVKGGALAIFAVLG 429

Qy         476 EPTEARV-VP-----------GRGICLDLAILDSAFLLSQV----------------AP 506
Db         430 IPLAITFSIPFALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPWDAMFGGGNLP 489

Qy         507 SLFMGSIVQLSQSVTAY 523
Db         490 AFVVGAVAATASAVLSF 506


RESULT 6
S38196
sucrose transport protein SUC2 - Arabidopsis thaliana
N;Alternate names: sucrose-proton symporter SUC2
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C;Accession: S38196; T00773
R;Sauer, N.K.
submitted to the EMBL Data Library, October 1993
A;Description: SUC1 and SUC2: two sucrose transporters from Arabidopsis thaliana.
A;Reference number: S38196
A;Accession: S38196
A;Molecule type: mRNA
A;Residues: 1-512 <SAU>
A;Cross-references: EMBL:X75382; NID:g407091; PID:g407092
R;Vysotskaia, V.S.; Schwartz, J.R.; Toriumi, M.; Yu, G.; Oji, O.; Kwan, A.; Liu, S.; Li
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, I
submitted to the EMBL Data Library, July 1998
A;Description: Arabidopsis thaliana chromosome 1 BAC T22J18 sequence, complete sequence
A;Reference number: Z14202
A;Accession: T00773
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-490,'A',492-512 <VYS>
```

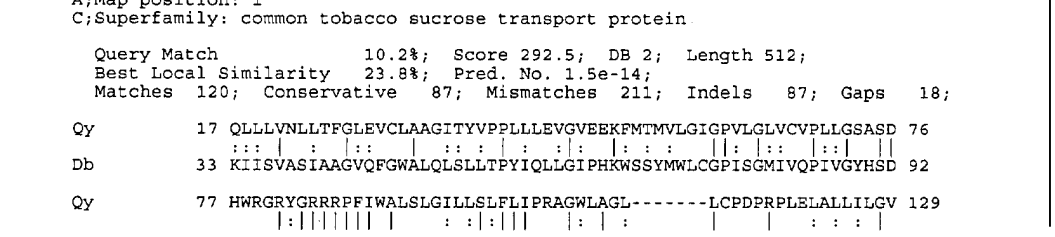
```
A;Cross-references: EMBL:AC003979; NID:g3172156; PID:g3287687; GSPDB:GN00059; ATSP:T22J1
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: SUC2; ATSP:T22J18.12
A;Map position: 1
A;Introns: 419/3; 441/1; 455/3
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: sugar transport

  Query Match             11.3%;  Score 323.5;  DB 2;  Length 512;
  Best Local Similarity   24.1%;  Pred. No. 6.6e-17;
  Matches  123;  Conservative   92;  Mismatches  195;  Indels  101;  Gaps   18;

Qy         14 RKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
              |  ::: |: :  |::   |  :: : | :  :|:  |: :::   ||: |::  |::|
Db         28 RLRKIISVSSIAAGVQFGWALQLSLLTPYVQLLGIPHKWASLIWLCGPISGMLVQPIVGY 87

Qy         74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLI 126
               ||    |:|||||||| |  :  : :::|||   |: | :       |   |:   :|:
Db         88 HSDRCTSRFGRRRPFIVAGAGLVTVAVFLI---GYAADIGHSMGDQLDKPPKTRAIAIFA 144

Qy        127 LGVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY-------LL 178
              ||  :||         |  | |:||   :    | | : ::| :::|  |||       |
Db        145 LGFWILDVANNTLQGPCRAFLADLSAGNAKKTRTANAFFSFFMAVGNVLGYAAGSYRNLY 204

Qy        179 PAIDWD-TSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
              : :  | :    |    : | |  :||: |     :|   :|    |   |:|: :
Db        205 KVVPFTMTESCDLYCANLKTCFFLSITLL-LIVTFVSLCYVKEKPWTPEPTADGKA---- 259

Qy        238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
                : :  |  :     :|     | :   :|:|   | || ||::|
Db        260 --------SNVPF--FGEIFGAFKEL----KRPMWMLLIVTALNWIAWFPFLLFDTDWMG 305

Qy        298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLA 357
              :| |    |     ::: |::||| |:||| |   :    || :: : :: |  |  |
Db        306 REVYGGNSDATATAASKKLYNDGVRAGALGLMLNAIVLGFMSLGVEWIGRKLG-GAKRLW 364

Qy        358 SVAAFPVAAGATCLSHSVAV--------------------VTASA----ALTG----FT 388
              :  | :   | ||: :| |                    |||  |    |: |      |
Db        365 GIVNFIL---AICLAMTVVVTKQAENHRRDHGGAKTGPPGNVTAGALTLFAILGIPQAIT 421

Qy        389 FSALQILPYTLASLYHREK-------------QVFLPKYRGDTGGASSEDSLMTSFLPGP 435
              ||    :|: |||::                 : :|:     ||   ::      :|
Db        422 FS----IPFALASIFSTNSGAGQGLSLGVLNLAIVVPQMVISVGGGPFDELFGGGNIPA- 476

Qy        436 KPGAPFPNGHVGAGGSGLL-------PPPPA 459
                |  | :  |  ||:|       ||| |
Db        477 -----FVLGAIAAAVSGVLGLTVLPSPPPDA 502


RESULT 7
G86360
probable sucrose-proton symporter SUC2 protein T22J18.12 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2001
C;Accession: G86360
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: G86360
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <STO>
A;Cross-references: GB:AE005172; NID:g3287687; PIDN:AAC25515.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: common tobacco sucrose transport protein

  Query Match             11.3%;  Score 323.5;  DB 2;  Length 512;
  Best Local Similarity   24.1%;  Pred. No. 6.6e-17;
  Matches  123;  Conservative   92;  Mismatches  195;  Indels  101;  Gaps   18;

Qy         14 RKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
              |  ::: |: :  |::   |  :: : | :  :|:  |: :::   ||: |::  |::|
Db         28 RLRKIISVSSIAAGVQFGWALQLSLLTPYVQLLGIPHKWASLIWLCGPISGMLVQPIVGY 87

Qy         74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLI 126
               ||    |:|||||||| |  :  : :::|||   |: | :       |   |:   :|:
Db         88 HSDRCTSRFGRRRPFIVAGAGLVTVAVFLI---GYAADIGHSMGDQLDKPPKTRAIAIFA 144

Qy        127 LGVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY-------LL 178
              ||  :||         |  | |:||   :    | | : ::| :::|  |||       |
Db        145 LGFWILDVANNTLQGPCRAFLADLSAGNAKKTRTANAFFSFFMAVGNVLGYAAGSYRNLY 204

Qy        179 PAIDWD-TSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
              : :  | :    |    : | |  :||: |     :|   :|    |   |:|: :
Db        205 KVVPFTMTESCDLYCANLKTCFFLSITLL-LIVTFVSLCYVKEKPWTPEPTADGKA---- 259

Qy        238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
                : :  |  :     :|     | :   :|:|   | || ||::|
Db        260 --------SNVPF--FGEIFGAFKEL----KRPMWMLLIVTALNWIAWFPFLLFDTDWMG 305

Qy        298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLA 357
              :| |    |     ::: |::||| |:||| |   :    || :: : :: |  |  |
Db        306 REVYGGNSDATATAASKKLYNDGVRAGALGLMLNAIVLGFMSLGVEWIGRKLG-GAKRLW 364

Qy        358 SVAAFPVAAGATCLSHSVAV--------------------VTASA----ALTG----FT 388
              :  | :   | ||: :| |                    |||  |    |: |      |
Db        365 GIVNFIL---AICLAMTVVVTKQAENHRRDHGGAKTGPPGNVTAGALTLFAILGIPQAIT 421

Qy        389 FSALQILPYTLASLYHREK-------------QVFLPKYRGDTGGASSEDSLMTSFLPGP 435
              ||    :|: |||::                 : :|:     ||   ::      :|
Db        422 FS----IPFALASIFSTNSGAGQGLSLGVLNLAIVVPQMVISVGGGPFDELFGGGNIPA- 476

Qy        436 KPGAPFPNGHVGAGGSGLL-------PPPPA 459
                |  | :  |  ||:|       ||| |
Db        477 -----FVLGAIAAAVSGVLGLTVLPSPPPDA 502


RESULT 8
S48789
sucrose transport protein - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 26-May-2000
C;Accession: S48789
R;Buerkle, X.Y.Z.; Frommer, W.B.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48787
A;Accession: S48789
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-507 <BUE>
A;Cross-references: EMBL:X82276; NID:g575350; PID:g575351
C;Superfamily: common tobacco sucrose transport protein

  Query Match             11.1%;  Score 317.5;  DB 2;  Length 507;
  Best Local Similarity   24.3%;  Pred. No. 1.8e-16;
  Matches  125;  Conservative   87;  Mismatches  203;  Indels   99;  Gaps   18;

Qy          4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
              :||          ::::|  :  |::   |  :: : | :  :|:  || : :   ||:
Db         23 KLW----------KIIMVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 72

Qy         64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDP-----R 118
              |::  |::|  ||:   |:|||| || | :  : :::|||  |  | |    ||     :
Db         73 GMIVQPVVGYYSDNCSSRFGRRRGFIAAGAALVTIAVFLIGFAADL-GHATGDPLGKGSK 131

Qy        119 PLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLL 178
```

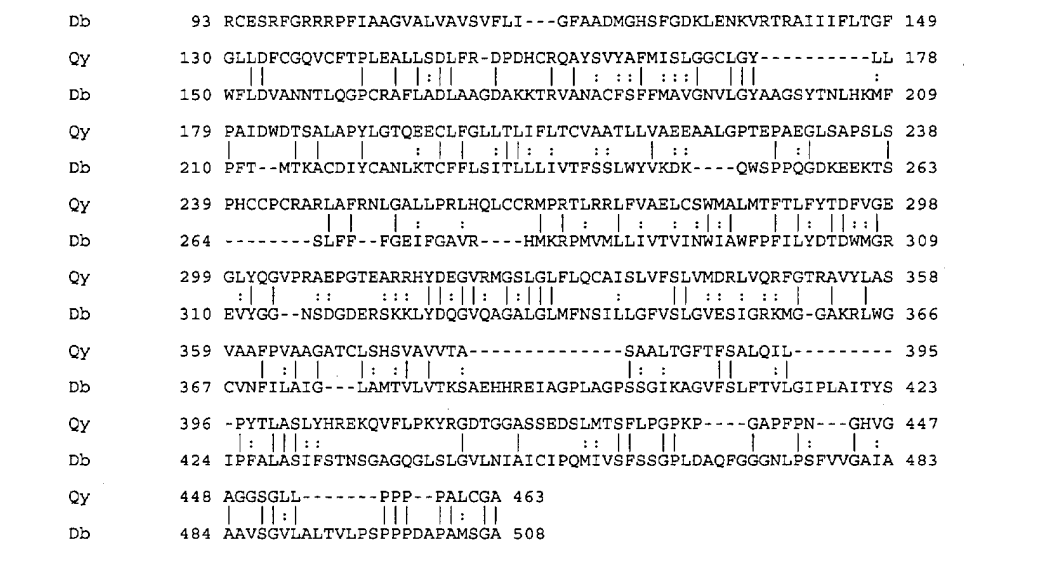
```
Db        132 PRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKARMRTSNAFFSFFMAVGNVLGYAA 191

Qy        179 PAID--------WDTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEEAALGPTEPA 229

Db        192 GSYSRLCKIFPFSKTPACDIYCANLKSCFFIAVFLLLSLTILALT--VVRENELPEKDEH 249

Qy        230 EGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFT 289

Db        250 E------IDEKAGARKSKVPF--FGEIFGALKDL----PRPMWILLLVTSLNWIARFPFF 297

Qy        290 LFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 349

Db        298 LYDTDWMAKEVYGG--KVGDG----RLYDLGVHAGALGLLLNSVVLGFMSLSVEFLGKKI 351

Qy        350 GTRAVYLASVAAFPVAAGATCLSHSVAVV-------------TASAALTGFTFSALQI-- 394

Db        352 G-GVKRLWGILNFVL---AVCMALTVLVTKMAEKSRQYDAHGTLMAPTSGVKIGALTLFA 407

Qy        395 -----------LPYTLASLYHREK-------------QVFLPKYRGDTGGASSEDSLMTS 430

Db        408 VLGIPLAVTFSVPFALASIFSSNAGSGQGLSLGVLNLAIVVPQMLVSIAGGPWDDLFGGG 467

Qy        431 FLPGPKPGAPFPNGHVGAGGSG-----LLPPPPA 459

Db        468 NLPG------FIVGAVAAAASGILALTMLPSPPA 495


RESULT 9
S43142
sucrose transport protein - castor bean
N;Alternate names: sucrose carrier
C;Species: Ricinus communis (castor bean)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 26-May-2000
C;Accession: S43142
R;Weig, A.; Komor, E.
submitted to the EMBL Data Library, March 1994
A;Description: A sucrose carrier from Ricinus communis.
A;Reference number: S43142
A;Accession: S43142
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-533 <WEI>
A;Cross-references: EMBL:Z31561; NID:g468561; PID:g468562
C;Superfamily: common tobacco sucrose transport protein

  Query Match             10.9%;  Score 311;  DB 2;  Length 533;
  Best Local Similarity   23.3%;  Pred. No. 6.1e-16;
  Matches  134;  Conservative   95;  Mismatches  225;  Indels  120;  Gaps   19;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76

Db         37 KVVMVASIAAGIQFGWALQLSLLTPYVQLLGIPHTWAAFIWLCGPISGMLVQPIVGYHSD 96

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRA---GWLAG-LLCPDPRPLELALLILGVGLL 132

Db         97 RCTSRFGRRRPFIASGAAFVAIAVFLIGYAADLGHLSGDSLDKSPKTRAIAIFVVGFWIL 156

Qy        133 DFCGQVCFTPLEALLSDLF-RDPDHCRQAYSVYAFMISLGGCLGYLLPAID--------W 183

Db        157 DVANNMLQGPCRALLADLSGTSQKKTRTANALFSFFMAVGNVLGYAAGAYTHLYKLFPFT 216

Qy        184 DTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGL--------SAP 235

Db        217 KTTACDVYCANLKSCFFISIVLLLSLTVLALSYVKEK----PWSPDQAVDNAEDDTASQA 272

Qy        236 SLSPHCCPCRARL--AFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293

Db        273 SSSAQPMPFFGEILGAFKNL--------------KRPMWILLLVTCLNWIAWFPFLLFDT 318

Qy        294 DFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRA 353

Db        319 DWMGREVYGG--DSSGSAEQLKLYDRGVRAGALGLMLNSVVLGFTSLGVEVLARGVG-GV 375

Qy        354 VYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPK 413

Db        376 KRLWGIVNFVL---AVCLAMTV-LVTKQAEST------------------------RR 405

Qy        414 YRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVV 473

Db        406 FATVSGGAK---------VP--------------------LPPP---SGVKAGALALFAV 433

Qy        474 VGEPTEARV-VP----------GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL--SQS 519

Db        434 MGVPQAITYSIPFALASIFSNTSGAGQGLSLGVLNLSIVIPQMIVSVAAGPWDALFGGGN 493

Qy        520 VTAYMVSAAGLGLVAIYFATQVVFDKSDLAKYSA 553

Db        494 LPAFVVGAVAALASGIFALTMLPSPQPDMPSAKA 527


RESULT 10
T14339
sucrose-proton transport protein - carrot
N;Alternate names: sucrose/H+ symporter protein
C;Species: Daucus carota (carrot)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T14339
R;Shakya, R.; Sturm, A.
Plant Physiol. 118, 1473-1480, 1998
A;Title: Characterization of source- and sink-specific sucrose/H+ symporters from carro
A;Reference number: Z17991; MUID:99063785; PMID:9847123
A;Accession: T14339
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-501 <SHA>
A;Cross-references: EMBL:Y16766; NID:g2969886; PIDN:CAA76367.1; PID:g2969887
A;Experimental source: cultivar Namtaise: leaf
C;Genetics:
A;Gene: SUT1a
C;Superfamily: common tobacco sucrose transport protein

  Query Match             10.7%;  Score 305.5;  DB 2;  Length 501;
  Best Local Similarity   21.6%;  Pred. No. 1.5e-15;
  Matches  124;  Conservative   91;  Mismatches  213;  Indels  145;  Gaps   14;

Qy         18 LLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDH 77

Db         30 LLRVASVACGIQFGWALQLSLLTPYVQELGIPHAWSSIIWLCGPLSGLLVQPIVGHMSDQ 89

Qy         78 WRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL--CPDPRPLELALLILGVGLLDFC 135

Db         90 CTSKYGRRRPFIVAGGTAIILAVIIIAHSADIGGLLGDTADNKTMAIVAFVIGFWILDVA 149

Qy        136 GQVCFTPLEALLSDLF-RDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGT 194

Db        150 NNMTQGPCRALLADLTGNDARRTRVANAYFSLFMAIGNVLGY----------ATGAYSGW 199

Qy        195 QEECLFGLLTLIFLTC-------------VAATLLVAEEAAL---------GPTEPAEGL 232

Db        200 YKVFPFSLTSSCTINCANLKSAFYIDIIFIIITTYISISAAKERPRISSQDGPQFSEDGT 259

Qy        233 SAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFY 292

Db        260 AQ---SGHI-----------EEAFLWELFGTFRLLPGSVWVILLVTCLNWIGWFPFILFD 305

Qy        293 TDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTR 352

Db        306 TDWMGREIYGGEP------NQGQSYSDGVRMGAFGLMMNSVVLGITSVLMEKLCRIWGSG 359

Qy        353 AVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLP 412

Db        360 FMW----------------------GLSNILMTICFFAMLLITFIAKNMDY-------- 388

Qy        413 KYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRV 472
```

```
Db        389 ----------------------------------------GTNPPPN---GIVISALIVFA 406

Qy        473 VVGEPTEARV-VP-----------GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL---- 516
              ::| |      ||           | |  | : :|: | :: ||  ||  |    ||
Db        407 ILGIPLAITYSVPYALVSTRIESLGLGQGLSMGVLNLAIVVPQVIVSLGSGPWDQLFGGG 466

Qy        517 -SQSVTAYMVSAAGLGLVAIYFATQVVFDKSDL 548
               | :     :||   ||:|:    :   ||| |
Db        467 NSPAFVVAALSAFAAGLIALIAIRRPRVDKSRL 499


RESULT 11
S48788
sucrose transport protein - tomato (fragment)
C;Species: Lycopersicon esculentum (tomato)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 09-Jun-2000
C;Accession: S48788
R;Buerkle, X.Y.Z.; Frommer, W.B.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48787
A;Accession: S48788
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-428 <BUE>
A;Cross-references: EMBL:X82275
C;Superfamily: common tobacco sucrose transport protein

  Query Match            10.4%;  Score 297.5;  DB 2;  Length 428;
  Best Local Similarity  25.8%;  Pred. No. 5.1e-15;
  Matches  114;  Conservative   82;  Mismatches  163;  Indels   83;  Gaps   18;

Qy          4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
              :||          ::::|  :  |::   |  :: : | :  :|:  :| : :   ||:
Db         25 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHRFASFIWLCGPIS 74

Qy         64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDP-----R 118
              |::  |::|  ||:   |:|||||| | :  : :::||  |  | |    ||     :
Db         75 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVTIAVFLIGFAADL-GHASGDPLGKGSK 133

Qy        119 PLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLGY- 176
              |  :|: ::|  :||    :   | |||:||         | | : ::|  ::|  |||
Db        134 PRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGKMRTANAFFSFFKAVGNILGYA 193

Qy        177 ------LLPAIDWD-TSALAPYLGTQEECLF-GLLTLIFLTCVAATLL----VAEEAALG 224
                   |     : | |   |    : | |  :  |: || :| ||:    : |:  |
Db        194 AGSYSRLFKVFPPSKTKACDMYRANLKSCFFIAIFLLLSLTTLALTLVRENELPEKEELE 253

Qy        225 PTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMA 284
              |  |  |          ::::  |  | :   |  |    || :  | :    :|:|
Db        254 IDEKLSGAG----------KSKVPF--FGEIFGALKDL----PRPMWILLLVTCLNWIA 296

Qy        285 LMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDR 344
              | |: ||:: : :: |    : |    : || ||| |:||| ||  :    || ::
Db        297 WFPPFLYDTDWMAKEVFGG----QVGD--AKLYDLGVRAGALGLLLQSVVLGFMSLGVEF 350

Qy        345 LVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT------- 388
              | :: |  |  |  :. | :   | ||: :: |           |:  | | |
Db        351 LGKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRGHDAAGTLMGPTPGVKIGA 406

Qy        389 ---FSALQI-------LPYTLA 400
                 |:|| |       :|: ||
Db        407 LLLFAALGIPLAVTFSIPPALA 428


RESULT 12
S38657
sucrose transport protein ptp1 - common plantain
N;Alternate names: sucrose transporter ptp1
C;Species: Plantago major (common plantain)
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 26-May-2000
C;Accession: S38657
R;Gahrtz, M.; Stolz, J.; Sauer, N.
submitted to the EMBL Data Library, November 1993
A;Reference number: S38657
A;Accession: S38657
A;Molecule type: mRNA
A;Residues: 1-510 <GAH>
A;Cross-references: EMBL:X75764; NID:g415987; PIDN:CAA53390.1; PID:g415988
C;Genetics:
A;Gene: ptp1
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: sugar transport

  Query Match            10.3%;  Score 294;  DB 2;  Length 510;
  Best Local Similarity  24.7%;  Pred. No. 1.1e-14;
  Matches  126;  Conservative   77;  Mismatches  199;  Indels  108;  Gaps   19;

Qy         18 LLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDH 77
              : ||  :  |::   |  :: : | :  :|:  |: : :   ||: |::  |::|  ||:
Db         29 IFLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWASYIWLCGPISGMIVQPVVGYFSDN 88

Qy         78 WRGRYGRRRPFIWALSLGIL-LSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
                |:|||||| |   |:: ::: ||   |: | |       |    :|  : : : |
Db         89 CTSRFGRRRPFI-AAGAGLVGVAVVLI---GFAADLGHAGGDSLGDGLKPRAIGVFVFGF 144

Qy        130 GLLDFCGQVCFTPLEALLSDLF-RDPDHCRQAYSVYAFMISLGGCLGYLLPAID------ 182
               :||    :   | |||:||   :       | | ::| :::|  |||   :
Db        145 WILDVANNMLQGPCRALLADLSGGNTKKMANANSFFSFFMAVGNVLGYAAGSYSRMYKVF 204

Qy        183 --WDTSALAPYLGTQEEC-LFGLLTLIFLTCVAATLL-----VAEEAALGPTEPAEGLSA 234
                 | |   |    : | :  :  || || :| :::     ||||      |   :|
Db        205 PFSKTKACDIYCANLKSCFIISITLLITLTTLALSIVREKRHVAEEQV---TAAKKGFKI 261

Qy        235 PSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTD 294
              |                  : | |      :|| :  | :    :|:|   | || ||
Db        262 P--------------------VFPELFGALKDLPRPMWVLLLVTALNWIAWFGFLLFDTD 301

Qy        295 FVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG---- 350
              ::|  :| |  :      |   |::||  |:||| |   :    || :  :  :  |
Db        302 WMGREVY-GETQQHKAPELAVIYNKGVSAGALGLMLNSIVLGFASLGVQYMARALGGVKR 360

Qy        351 -------TRAVYLA-SVAAFPVAAGATCLSHSVAVVTASAALTG--FTFSALQI------ 394
                    |: |   ::    ||:     |: |     |:   |    |||| |
Db        361 LWGVVNFILAICLCMTIVITKVASHHRPYSNGVLQTPESSVKIGALVVFSALGIPLAITF 420

Qy        395 -LPYTLASLYHREKQVFLPKYRGDTGGASSEDSL-----------MTSFLPGPKPG--- 438
               :|: |||:|               | |:    ||           : |   ||
Db        421 SVPFALASIY-----------STTTGSGQGLSLGVLNLAIVIPQMIVSVASGPWDAMFG 468

Qy        439 ----APFPNGHVGAGGSG-----LLPPPPA 459
                     |  | | |   ||      :|| |||
Db        469 GGNLPAFVVGAVAAAASGIFAFTMLPSPPA 498


RESULT 13
S51114
sucrose-proton symporter - beet
C;Species: Beta vulgaris (beet)
C;Date: 07-May-1995 #sequence_revision 01-Sep-1995 #text_change 26-May-2000
C;Accession: S51114
R;Westram, A.; Eckhardt, U.; Frommer, W.B.; Riesmeier, J.W.
submitted to the EMBL Data Library, January 1995
A;Description: Sequence of a sugar beet sucrose transporter cDNA.
A;Reference number: S51114
A;Accession: S51114
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-523 <WES>
A;Cross-references: EMBL:X83850; NID:g633171; PID:g633172
C;Superfamily: common tobacco sucrose transport protein

  Query Match            10.3%;  Score 294;  DB 2;  Length 523;
  Best Local Similarity  24.3%;  Pred. No. 1.2e-14;
  Matches  125;  Conservative   78;  Mismatches  199;  Indels  112;  Gaps   20;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
```

RESULT 14
F96741
probable sucrose transport protein F17M19.4 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: F96741
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F96741
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <STO>
A;Cross-references: GB:AE005173; NID:g6978914; PIDN:AAF34306.1; GSPDB:GN00141
C;Genetics:
A;Gene: F17M19.4
A;Map position: 1
C;Superfamily: common tobacco sucrose transport protein

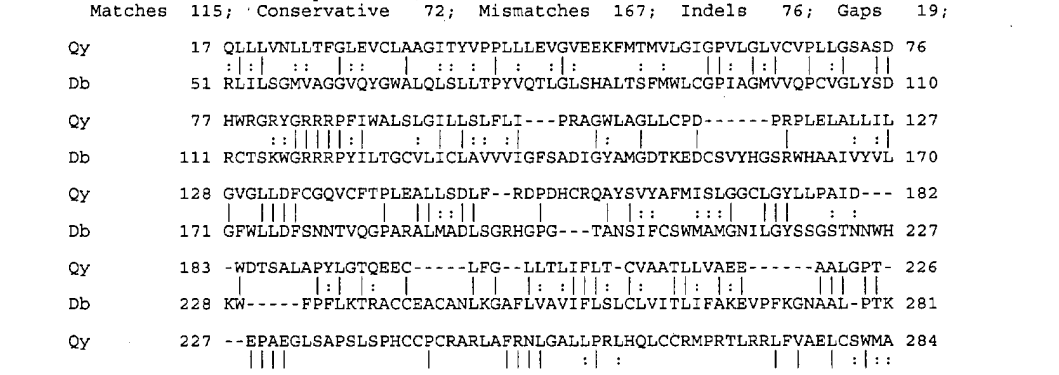
Query Match 10.2%; Score 292.5; DB 2; Length 512;
Best Local Similarity 23.8%; Pred. No. 1.5e-14;
Matches 120; Conservative 87; Mismatches 211; Indels 87; Gaps 18;





RESULT 15
T02982
probable sucrose transport protein - rice
C;Species: Oryza sativa (rice)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C;Accession: T02982
R;Hirose, T.; Imaizumi, N.; Scofield, G.N.; Furbank, R.T.; Ohsugi, R.
Plant Cell Physiol. 38, 1389-1396, 1997
A;Title: cDNA cloning and tissue specific expression of a gene for sucrose transporter
A;Reference number: Z14809; MUID:98182940; PMID:9522469
A;Accession: T02982
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-537 <HIR>
A;Cross-references: EMBL:D87819; NID:g2723470; PIDN:BAA24071.1; PID:g2723471
A;Experimental source: cultivar Nipponbare, leaf
C;Genetics:
A;Gene: SUT1
C;Superfamily: common tobacco sucrose transport protein

Query Match 10.2%; Score 292; DB 2; Length 537;
Best Local Similarity 26.7%; Pred. No. 1.7e-14;
Matches 115; Conservative 72; Mismatches 167; Indels 76; Gaps 19;



```
Db        282 SNEPAEPEGTG-------PLAVLKGFRNLPTGMPSV------------LIVTGL-TWLS 320

Qy        285 LMTFTLFYTDFVGEGLYQGVPR-AEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMD 343
                | |: ||::|  :| | |:  :|  ||   :::||| |: || |   :    | :::
Db        321 WFPFILYDTDWMGREIYHGDPKGTDPQIEA---FNQGVRAGAFGLLLNSIVLGFSSFLIE 377

Qy        344 RLVQRFGTRAVYLASVAAFPVAAGATCL--------SHSVA--VVTASAALTGFTFSALQ 393
               : :: | | |:: |     :|  || |         |      :||  ::
Db        378 PMCRKVGPRVVWVTSNFLVCIAMAATALISFWSLKDFHGTVQKAITADKSIKAVCLVLFA 437

Qy        394 ILPYTLASLY 403
                |   || ||
Db        438 FLGVPLAVLY 447
```

Search completed: August 25, 2004, 02:28:38
Job time : 51.1993 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 25, 2004, 01:58:37 ; Search time 29.5196 Seconds
(without alignments)
975.448 Million cell updates/sec

Title: US-09-593-793A-113
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	489.5	17.1	530	1	MATP_MOUSE	P58355 mus musculu
2	488.5	17.1	530	1	MATP_HUMAN	Q9umx9 homo sapien
3	451	15.8	751	1	PAST_MOUSE	Q8biv7 mus musculu
4	446	15.6	751	1	PAST_RAT	Q8k4s3 rattus norv
5	324.5	11.3	525	1	SUT_SPIOL	Q03411 spinacia ol
6	267	9.3	509	1	PAST_HUMAN	Q9y2w3 homo sapien
7	235.5	8.2	553	1	SUT1_SCHPO	O14091 schizosacch
8	142	5.0	544	1	YD74_SYNY3	P74168 synechocyst
9	136	4.8	541	1	GT10_HUMAN	O95528 homo sapien
10	125.5	4.4	399	1	TCR1_ECOLI	P02982 escherichia
11	121.5	4.2	503	1	PUR8_STRLP	P42670 streptomyce
12	119.5	4.2	491	1	AMPG_ECOLI	P36670 escherichia
13	117	4.1	440	1	YHJE_ECOLI	P37643 escherichia
14	117	4.1	495	1	SMVA_SALTY	P37594 salmonella
15	111.5	3.9	640	1	Y051_MYCTU	Q10880 mycobacteri
16	111	3.9	473	1	PHDK_NOCSK	O24723 nocardioide
17	110.5	3.9	680	1	CA1A_HUMAN	Q03692 homo sapien
18	110	3.8	368	1	GALT_HUMAN	O60755 homo sapien
19	110	3.8	606	1	NQOC_THETH	Q56227 thermus the
20	109.5	3.8	654	1	SPH2_HUMAN	Q9nra0 homo sapien
21	109	3.8	477	1	YPUM_RHOCA	P26176 rhodobacter
22	108.5	3.8	702	1	OPGB_XANCP	Q8pdd7 xanthomonas
23	107	3.7	419	1	CMLA_PSEAE	P32482 pseudomonas
24	105	3.7	448	1	YJ94_YEAST	P47159 saccharomyc
25	104	3.6	476	1	MELB_SALTY	P30878 salmonella
26	103	3.6	465	1	FTSW_MYCLE	Q50186 mycobacteri
27	102.5	3.6	461	1	PUCC_RHOCA	P23462 rhodobacter
28	102	3.6	438	1	SHIA_ECOLI	P76350 escherichia
29	102	3.6	702	1	HPPA_RHORU	O68460 rhodospiril
30	102	3.6	794	1	HPPA_STRCO	Q9x913 streptomyce
31	102	3.6	801	1	HPPA_STRAW	Q82ej8 streptomyce
32	101.5	3.5	428	1	YXIO_BACSU	P42306 bacillus su
33	101.5	3.5	510	1	NANT_YERPE	Q8zch3 yersinia pe
34	101	3.5	437	1	BRAZ_PSEAE	P25185 pseudomonas
35	101	3.5	471	1	MELB_ENTAE	O07366 enterobacte
36	101	3.5	471	1	MELB_KLEPN	Q02581 klebsiella
37	100	3.5	404	1	Y4XM_RHISN	P55705 rhizobium s
38	100	3.5	473	1	YIHO_SALTY	Q9i7r4 salmonella
39	99	3.5	379	1	CYB_SORHY	O79453 sorex hoyi
40	99	3.5	379	1	CYB_SORJA	Q8se00 sorex jacks
41	99	3.5	386	1	CV03_HUMAN	Q9y3p4 homo sapien
42	99	3.5	481	1	LMRA_STRLN	P46104 streptomyce
43	98.5	3.4	613	1	NUOL_ECOLI	P33607 escherichia
44	98.5	3.4	685	1	FHUB_SALTY	O87656 salmonella
45	98	3.4	379	1	CYB_SORCI	O79451 sorex ciner

ALIGNMENTS

```
RESULT 1
MATP_MOUSE
ID   MATP_MOUSE     STANDARD;      PRT;   530 AA.
AC   P58355;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Membrane-associated transporter protein (AIM-1 protein) (Melanoma
DE   antigen AIM1) (Underwhite protein).
GN   MATP OR AIM1 OR UW.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ddY; TISSUE=Eye, Kidney, and Uterus;
RX   MEDLINE=21372467; PubMed=11479596;
RA   Fukamachi S., Shimada A., Shima A.;
RT   "Mutations in the gene encoding B, a novel transporter protein, reduce
RT   melanin content in medaka.";
RL   Nat. Genet. 28:381-385(2001).
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANTS UW-DBR ASN-153 AND PRO-435.
RX   MEDLINE=21473748; PubMed=11574907;
RA   Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davisson M.T.,
RA   King R.A., Brilliant M.H.;
RT   "Mutations in the human orthologue of the mouse underwhite gene (uw)
RT   underlie a new form of oculocutaneous albinism, OCA4.";
RL   Am. J. Hum. Genet. 69:981-988(2001).
CC   -!- FUNCTION: Melanocyte differentiation antigen. May transport
CC       substances required for melanin biosynthesis (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By
CC       similarity).
CC   -!- TISSUE SPECIFICITY: Melanocytes, eyes, kidney and uterus.
CC   -!- DISEASE: Defects in MATP are the cause of the UW-dbr phenotype
CC       that results in loss of nearly all pigmentation in the homozygous
CC       state.
CC   -!- SIMILARITY: Belongs to the glycoside-pentoside-hexuronide (GPH)
CC       cation symporter transporter (TC 2.A.2) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF360357; AAK81713.1; -.
DR   MGD; MGI:2153040; Matp.
KW   Melanin biosynthesis; Transmembrane; Antigen; Glycoprotein; Vision;
KW   Disease mutation; Albinism.
FT   DOMAIN        1     45       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     46     66       1 (POTENTIAL).
FT   DOMAIN       67     68       EXTRACELLULAR (POTENTIAL).
```

```
FT   TRANSMEM      69     89       2 (POTENTIAL).
FT   DOMAIN        90    105       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     106    126       3 (POTENTIAL).
FT   DOMAIN       127    138       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     139    159       4 (POTENTIAL).
FT   DOMAIN       160    184       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     185    205       5 (POTENTIAL).
FT   DOMAIN       206    216       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     217    237       6 (POTENTIAL).
FT   DOMAIN       238    318       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     319    339       7 (POTENTIAL).
FT   DOMAIN       340    366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     367    387       8 (POTENTIAL).
FT   DOMAIN       388    398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     399    419       9 (POTENTIAL).
FT   DOMAIN       420    425       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     426    446       10 (POTENTIAL).
FT   DOMAIN       447    477       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     478    498       11 (POTENTIAL).
FT   DOMAIN       499    504       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     505    525       12 (POTENTIAL).
FT   DOMAIN       526    530       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT      153    153       D -> N (IN UW-DBR).
FT   VARIANT      435    435       S -> P (IN UW-DBR).
SQ   SEQUENCE   530 AA;  57961 MW;  F4EDEAD07916D9FC CRC64;

  Query Match           17.1%;  Score 489.5;  DB 1;  Length 530;
  Best Local Similarity   26.1%;  Pred. No. 1.9e-28;
  Matches  146;  Conservative   85;  Mismatches  228;  Indels  101;  Gaps   12;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
              :|:: ::  || | | |    || |:|| ||: :   :||  : |:|| :  |::|||||
Db         34 RLVMHSMAMFGREFCYAVEAAYVTPVLLSVGLPKSLYSMVWLLSPILGFLLQPVVGSASD 93

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRP---LELALLILGVGLLD 133
              | | |:|||||:|  |:: :|| : |      :   |  :||      ::: ::|| | |
Db         94 HCRARWGRRRPYILTLAIMMLLGMALYLNGDAVVSALVANPRQKLIWAISITMVGVVLFD 153

Qy        134 FCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLG 193
              |      |::| | |:    |  :    :|     || |||:| ||||    |   ||
Db        154 FSADFIDGPIKAYLFDVCSHQDK-EKGLHYHALFTGFGGALGYILGAIDWVHLDLGRLLG 212

Qy        194 TQEECLFGLLTLIFLTCVAATLLVAEEAAL--GPTEP-------AEGLSAPSLSPHCCPC 244
              |: : :|    |: : |    |   || |    |:|          |||  :  :
Db        213 TEFQVMFFFSALVLILCFITHLCSIPEAPLRDAATDPPSQQDPQGSSLSASGMHEY---- 268

Qy        245 RARLAFRNLGA----------------------LLPRLHQLCCRMPRTLRRLFVAELCSW 282
              :    :| ||                     :  | :    ||   | | |: |  |
Db        269 GSIEKVKNGGADTEQPVQEWKNKKPSGQSQRTMSMKSLLRALVNMPSHYRCLCVSHLIGW 328

Qy        283 MALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVM 342
              | ::  ||:|||:|: : | | |      :     |: || :|  || :    | |:|
Db        329 TAFLSNMLFFTDFMGQIVYHGDPYGAHNSTEFLIYERGVEVGCWGLCINSVFSSVYSYFQ 388

Qy        343 DRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASL 402
              :|   | : :|       | :  |   | :|       :: |   | |  :|: | :
Db        389 KAMVSYIGLKGLYFMGYLLFGLGTGFIGLFPNVYSTLVLCSMFGVMSSTLYTVPFNLIAE 448

Qy        403 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCG 462
              ||||::    | :|          || |   |      | |:         |
Db        449 YHREEE----KEKGQEA----------------PGGPDNQGR----GKGV------DCA 477

Qy        463 ASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTA 522
              |  | | :  .      |::: | |                    :| :| :: ||
Db        478 ALTCMVQL---------AQILVGGG---------------------LGFLVNMAGSVVV 506

Qy        523 YMVSAAGLGLVAIYFATQVV 542
              :::|: : |:   |    |
Db        507 VVITASAVSLIGCCFVALFV 526


RESULT 2
```

```
MATP_HUMAN
ID   MATP_HUMAN     STANDARD;      PRT;   530 AA.
AC   Q9UMX9; Q9BTM3;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Membrane-associated transporter protein (AIM-1 protein) (Melanoma
DE   antigen AIM1).
GN   MATP OR AIM1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM AIM-1A).
RC   TISSUE=Melanoma;
RX   MEDLINE=21115844; PubMed=11221837;
RA   Harada M., Li Y.F., El-Gamil M., Rosenberg S.A., Robbins P.F.;
RT   "Use of an in vitro immunoselected tumor line to identify shared
RT   melanoma antigens recognized by HLA-A*0201-restricted T cells.";
RL   Cancer Res. 61:1089-1094(2001).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM AIM-1B).
RC   TISSUE=Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   ALTERNATIVE SPLICING.
RA   Ferro S.;
RL   Unpublished observations (NOV-2001).
RN   [4]
RP   DISEASE, AND VARIANT LEU-374.
RX   MEDLINE=21473748; PubMed=11574907;
RA   Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davisson M.T.,
RA   King R.A., Brilliant M.H.;
RT   "Mutations in the human orthologue of the mouse underwhite gene (uw)
RT   underlie a new form of oculocutaneous albinism, OCA4.";
RL   Am. J. Hum. Genet. 69:981-988(2001).
CC   -!- FUNCTION: Melanocyte differentiation antigen. May transport
CC       substances required for melanin biosynthesis (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By
CC       similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC         Comment=Additional isoforms seem to exist;
CC       Name=AIM-1a;
CC         IsoId=Q9UMX9-1; Sequence=Displayed;
CC       Name=AIM-1b;
CC         IsoId=Q9UMX9-2; Sequence=VSP_006297, VSP_006298, VSP_006299;
CC       Name=AIM-1c;
CC         IsoId=Q9UMX9-3; Sequence=VSP_006296;
CC   -!- TISSUE SPECIFICITY: Expressed in most melanoma cell lines and
CC       melanocytes.
CC   -!- DISEASE: Defects in MATP are the cause of oculocutaneous albinism
```

```
CC       type 4 (OCA4) [MIM:606574]. OCA4 is characterized by
CC       hypopigmentation of skin, hair and eyes. It leads to reduced
CC       visual acuity.
CC   -!- SIMILARITY: Belongs to the glycoside-pentoside-hexuronide (GPH)
CC       cation symporter transporter (TC 2.A.2) family.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 188.
CC   -!- CAUTION: The described alternatively spliced isoforms are inferred
CC       using information from ESTs.
CC   -!- DATABASE: NAME=Mutations of the MATP gene;
CC       NOTE=Retina International's Scientific Newsletter;
CC       WWW="http://www.retina-international.com/sci-news/matpmut.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF172849; AAD51812.1; -.
DR   EMBL; BC003597; AAH03597.1; ALT_FRAME.
DR   Genew; HGNC:16472; MATP.
DR   MIM; 606202; -.
DR   MIM; 606574; -.
KW   Melanin biosynthesis; Transmembrane; Antigen; Glycoprotein; Vision;
KW   Polymorphism; Albinism; Alternative splicing.
FT   DOMAIN        1     46       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     47     67       1 (POTENTIAL).
FT   DOMAIN       68     68       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     69     89       2 (POTENTIAL).
FT   DOMAIN       90    110       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    111    131       3 (POTENTIAL).
FT   DOMAIN      132    138       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    139    159       4 (POTENTIAL).
FT   DOMAIN      160    184       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    185    205       5 (POTENTIAL).
FT   DOMAIN      206    216       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    217    237       6 (POTENTIAL).
FT   DOMAIN      238    318       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    319    339       7 (POTENTIAL).
FT   DOMAIN      340    366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    367    387       8 (POTENTIAL).
FT   DOMAIN      388    398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    399    419       9 (POTENTIAL).
FT   DOMAIN      420    425       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    426    446       10 (POTENTIAL).
FT   DOMAIN      447    477       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    478    498       11 (POTENTIAL).
FT   DOMAIN      499    504       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    505    525       12 (POTENTIAL).
FT   DOMAIN      526    530       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    129    187       Missing (in isoform AIM-1c).
FT                                /FTId=VSP_006296.
FT   VARSPLIC    188    295       Missing (in isoform AIM-1b).
FT                                /FTId=VSP_006297.
FT   VARSPLIC    386    406       YFQKVLVSYIGLKGLYFTGYL -> CKSFSLLRMSSKSFWS
FT                                STTWI (in isoform AIM-1b).
FT                                /FTId=VSP_006298.
FT   VARSPLIC    407    530       Missing (in isoform AIM-1b).
FT                                /FTId=VSP_006299.
FT   VARIANT     374    374       F -> L.
FT                                /FTId=VAR_012162.
SQ   SEQUENCE   530 AA;  58301 MW;  F14A4BACAA8FF31B CRC64;

  Query Match             17.1%;  Score 488.5;  DB 1;  Length 530;
  Best Local Similarity   26.4%;  Pred. No. 2.3e-28;
  Matches  150;  Conservative   78;  Mismatches  233;  Indels  107;  Gaps   12;

Qy          8 SRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVC 67
              |||: |  |          || | | |    || |:|| ||:     ::|  : |:|| :

Db         33 SRLIMHSMAM--------FGREFCYAVEAAYVTPVLLSVGLPSSLYSIVWFLSPILGFLL 84

Qy         68 VPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPR---PLELAL 124
              |::||||||  | |:|||||:|   | : :|: : |       :   |  :||      :::
Db         85 QPVVGSASDHCRSRWGRRRPYILTLGVMMLVGMALYLNGATVVAALIANPRRKLVWAISV 144

Qy        125 LILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWD 184
              ::|| | ||        |::| | |:    |  :    :|      || ||||| ||||
Db        145 TMIGVVLFDFAADFIDGPIKAYLFDVCSHQDK-EKGLHYHALFTGFGGALGYLLGAIDWA 203

Qy        185 TSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPC 244
                |   |||: : :|    |:  |    |   || | || |:|:  |  :|   |
Db        204 HLELGRLLGTEFQVMFFFSALVLTLCFTVHLCSISEAPL--TEVAKGI-PPQQTPQDPPL 260

Qy        245 RA------------------RLAFRNL-----------GALLPRLHQLCCRMPRTLRRL 274
              :                    || :                |  | :    ||   | |
Db        261 SSDGMYEYGSIEKVKNGYVNPELAMQGAKNKNHAEQTRRAMTLKSLLRALVNMPPHYRYL 320

Qy        275 FVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAI 334
              :: |  | | ::  ||:|||:|: :|:| | :   :      |: || :|  |  :
Db        321 CISHLIGWTAFLSNMLFFTDFMGQIVYRGDPYSAHNSTEFLIYERGVEVGCWGFCINSVF 380

Qy        335 SLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQI 394
              | ::|     ||   | : :|      | :  |   |  :|       :| |   | |
Db        381 SSLYSYFQKVLVSYIGLKGLYFTGYLLFGLGTGFIGLFPNVYSTLVLCSLFGVMSSTLYT 440

Qy        395 LPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLL 454
              :|: | : ||||::    | |                 ||  |
Db        441 VPFNLITEYHREEE----KERQQA-------------PGGDP----------------- 465

Qy        455 PPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIV 514
                        | |||             |:|  :| | |     |:|:     :| :|
Db        466 -----------DNSVR-------------GKG--MDCATLTCMVQLAQILVGGGLGFLV 498

Qy        515 QLSQSVTAYMVSAAGLGLVAIYFATQVV 542
              : :|   :::|: : |:    |     |
Db        499 NTAGTVVVVVITASAVALIGCCFVALFV 526


RESULT 3
PAST_MOUSE
ID   PAST_MOUSE     STANDARD;      PRT;   751 AA.
AC   Q8BIV7;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Proton-associated sugar transporter A (PAST-A) (Deleted in
DE   neuroblastoma 5 protein homolog) (DNb-5 homolog).
GN   DNB5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
```

```
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- FUNCTION: Mediates glucose uptake along the pH gradient (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Belongs to the glycoside-pentoside-hexuronide (GPH)
CC       cation symporter transporter (TC 2.A.2) family.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC       frameshift in position 217.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AK082651; BAC38560.1; ALT_FRAME.
DR   PIR; PT0633; PT0633.
DR   MGD; MGI:2653235; Dnb5.
KW   Transmembrane; Transport; Sugar transport; Symport.
FT   TRANSMEM     93    113       POTENTIAL.
FT   TRANSMEM    123    143       POTENTIAL.
FT   TRANSMEM    155    175       POTENTIAL.
FT   TRANSMEM    191    211       POTENTIAL.
FT   TRANSMEM    233    253       POTENTIAL.
FT   TRANSMEM    268    288       POTENTIAL.
FT   TRANSMEM    536    556       POTENTIAL.
FT   TRANSMEM    576    596       POTENTIAL.
FT   TRANSMEM    606    626       POTENTIAL.
FT   TRANSMEM    630    650       POTENTIAL.
FT   TRANSMEM    688    708       POTENTIAL.
FT   TRANSMEM    710    730       POTENTIAL.
SQ   SEQUENCE   751 AA;  81420 MW;  8242480E154F6BE1 CRC64;

  Query Match           15.8%;  Score 451;  DB 1;  Length 751;
  Best Local Similarity   21.7%;  Pred. No. 1.8e-25;
  Matches  155;  Conservative   80;  Mismatches  212;  Indels  268;  Gaps   13;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
              :||    : ||:|   |    || |:||::|: ::  ::|  | |:|| :  ||||: ||
Db         86 ELLFNGCILFGIEFSYAMETAYVTPVLLQMGLPDQLYSLVWFISPILGFLLQPLLGAWSD 145

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDP---RPLELALLILGVGLLD 133
                |:|||||||  |::| || | |:    |   |:   |        : | : || |:|
Db        146 RCTSRFGRRRPFILVLAIGALLGLSLLLN-GRDIGMALADTATNHKWGILLTVCGVVLMD 204

Qy        134 FCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLG 193
              |       |  | : |:    |  |   :::| |  |||  ||::  | || ::    ||
Db        205 FSADSADNPSHAXMMDVCGPVDQDR-GLNIHALMAGLGGGFGYVVGGIHWDKTSFGRALG 263

Qy        194 TQEECLFGLLTLIFLTCVAATLLVAEEAALGPT-EPAEGLSAPSL--------------- 237
              |   ::   :       ||:  |  | |  |    : :|||
Db        264 GQLRVIYVFTAITLSVTTVLTLISIPERPLRPLGEKRTAMKSPSLPLPPSPPVLLEEGAG 323

Qy        238 ----------------SPHCCP------------------------------------- 243
                             ||  |
Db        324 DALPSTTATSLYASFSSPISPPSPLTPKYGSFISRDSSLTGINEFASSFGTSNIDSVLID 383

Qy        244 ---------------------------------CRARLAFRNLGALL------------- 257
                                               |: |   |  | |

Db        384 CFTAGHDNYLALPSSVPRQAISVSFPRAPDGFYCQERGLERREGPLTLGSDGDVLRVGSL 443

Qy        258 ----PR---------------------------------------------------- 259
                  ||
Db        444 DTSKPRASGILKRPQTLALPDVAGGNGPETSRRRNVTFSQQVANILLNGVKYESELTGSS 503

Qy        260 --------LHQLCC---RMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAE 308
                     |   ||      ||: || | |      |::      ||||||:|| ::|| |:|
Db        504 EQSEQPLSLRHLCSTIYNMPKALRNLCVNHFLGWLSFEGMLLFYTDFMGEVVFQGDPKAP 563

Qy        309 PGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGA 368
              :|| : |: || ||  |: :    :  :| ::::| :    | :|  :  || :  |
Db        564 HTSEAYQKYNSGVTMGCWGMCIYAFSAAFYSAILEKLEECLSVRTLYFIAYLAFGLGTGL 623

Qy        369 TCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLM 428
              || :: || :     |  || |  |||:|   |:: |     |: |     || |
Db        624 ATLSRNLYVVLSLCTTYGILFSTLCTLPYSLLCDYYQSK-----KFAG-----SSADGTR 673

Qy        429 TSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEARVVPGRGI 488
                                                                      ||:
Db        674 ------------------------------------------------------RGM 676

Qy        489 CLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVF 543
                :|:::|   : |:|:  || :|           : || |     :||:: | |
Db        677 GVDISLLSCQYFLAQILVSLVLGP----------LTSAVGSANGVMYFSSLVSF 720


RESULT 4
PAST_RAT
ID   PAST_RAT       STANDARD;      PRT;   751 AA.
AC   Q8K4S3;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Proton-associated sugar transporter A (PAST-A).
GN   DNB5.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22305182; PubMed=12417639;
RA   Shimokawa N., Okada J., Haglund K., Dikic I., Koibuchi N., Miura M.;
RT   "Past-A, a novel proton-associated sugar transporter, regulates
RT   glucose homeostasis in the brain.";
RL   J. Neurosci. 22:9160-9165(2002).
CC   -!- FUNCTION: Mediates glucose uptake along the pH gradient.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: Predominantly expressed in brain.
CC   -!- SIMILARITY: Belongs to the glycoside-pentoside-hexuronide (GPH)
CC       cation symporter transporter (TC 2.A.2) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB075229; BAB97313.1; -.
KW   Transmembrane; Transport; Sugar transport; Symport.
FT   TRANSMEM     93    113       POTENTIAL.
FT   TRANSMEM    123    143       POTENTIAL.
FT   TRANSMEM    155    175       POTENTIAL.
FT   TRANSMEM    191    211       POTENTIAL.
FT   TRANSMEM    233    253       POTENTIAL.
FT   TRANSMEM    268    288       POTENTIAL.
FT   TRANSMEM    536    556       POTENTIAL.
FT   TRANSMEM    576    596       POTENTIAL.
FT   TRANSMEM    606    626       POTENTIAL.
```

```
FT   TRANSMEM     630    650       POTENTIAL.
FT   TRANSMEM     688    708       POTENTIAL.
FT   TRANSMEM     710    730       POTENTIAL.
SQ   SEQUENCE   751 AA;  81751 MW;  5C33E1C2E4A1B2EC CRC64;

  Query Match             15.6%;  Score 446; DB 1;  Length 751;
  Best Local Similarity   22.3%;  Pred. No. 4.1e-25;
  Matches  160;  Conservative   79;  Mismatches  206;  Indels  272;  Gaps   15;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
              :||    : ||:|    |     || |:||::|: ::  ::|  | |:|| :  |||:: ||
Db         86 ELLFNGCILFGIEFSYAMETAYVTPVLLQMGLPDQLYSLVWFISPILGFLLQPLLGAWSD 145

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDP---RPLELALLILGVGLLD 133
                |:|||||||  |::| || | |:   |   |:   |        : | : || |:|
Db        146 RCTSRFGRRRPFILVLAIGALLGLSLLLN-GRDIGMALADTATNHKWGILLTVCGVVLMD 204

Qy        134 FCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLG 193
              |      | | : |:    |  |   :::| |  |||  ||::  | || ::    ||
Db        205 FSADSADNPSHAYMMDVCGPVDQDR-GLNIHALMAGLGGGFGYVVGGIHWDKTSFGRALG 263

Qy        194 TQEECLFGLLTLIFLTCVAATLLVA------------------------------EEAA 222
              |   :: : | | |:       ||:                              || |
Db        264 GQLRVIY-IFTAITLSVTTVFTLVSIPERPLRPLGEKRTAMKSPSLPLPPSPPVLLEEGA 322

Qy        223 LGPTEP---AEGLSAPSLSPHCCP---------------------------------- 243
              | | |   | | | |   ||   |
Db        323 -GDTLPSTTATSLYASFSSPISPPSPLTPKYGSFISRDSSLTGINEFASSFGTSNIDSVL 381

Qy        244 ----------------------------------CRARLAFRNLGALL----------- 257
                                               |: |   |  | |
Db        382 IDCFTAGHDNYLALPSSVPRQAISVSFPRAPDGFYCQERGLERREGPLTLGLDGDVLRVG 441

Qy        258 ------PR---------------------------------------------------- 259
                    ||
Db        442 SLDTSKPRASGILKRPQTLALPDVAGGNGPETSRRRNVTFSQQVANILLNGVKYESELTG 501

Qy        260 ----------LHQLCC---RMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPR 306
                       | :||     ||| :| | |     |:     ||||||:|| ::|| |:
Db        502 SSEQSEQPLSLRRLCSTIYNMPRPVRNLCVNHFLGWLSFEGMLLFYTDFMGEVVFQGDPK 561

Qy        307 AEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAA 366
              |   :|| : |: || ||   |: :     :  :| ::::| :     | :|   :   | :
Db        562 APHASEAYQKYNSGVTMGCWGMCIYAFSAAFYSAILEKLEECLSVRTLYFIAYLLFGLGT 621

Qy        367 GATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDS 426
              |   || :: || :      |  || |  |||:|   |:: |     |: |     || |
Db        622 GLATLSRNLYVVLSLCTHYGILFSTLCTLPYSLLCDYYQSK-----KPAG-----SSADG 671

Qy        427 LMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEARVVPGR 486
                                                                        |
Db        672 TR-------------------------------------------------------R 674

Qy        487 GICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVF 543
              |: :|:::|   : |:|:  || :|           : || |     :|||: | |
Db        675 GMGVDISLLSCQYFLAQILVSLVLGP----------LTSAVGSANGVMYFASLVSF 720


RESULT 5
SUT_SPIOL
ID   SUT_SPIOL      STANDARD;      PRT;   525 AA.
AC   Q03411;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Sucrose transport protein (Sucrose permease) (Sucrose-proton
DE   symporter).
OS   Spinacia oleracea (Spinach).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Caryophyllales; Amaranthaceae; Spinacia.
OX   NCBI_TaxID=3562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leaf;
RX   MEDLINE=93099843; PubMed=1464305;
RA   Riesmeier J.W., Willmitzer L., Frommer W.B.;
RT   "Isolation and characterization of a sucrose carrier cDNA from
RT   spinach by functional expression in yeast.";
RL   EMBO J. 11:4705-4713(1992).
CC   -!- FUNCTION: Responsible for the transport of sucrose into the
CC       cell, with the concomitant export of a proton (symport system).
CC       Can also transport maltose at a lesser rate.
CC   -!- PATHWAY: Sucrose metabolism.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the glycoside-pentoside-hexuronide (GPH)
CC       cation symporter transporter (TC 2.A.2) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67125; CAA47604.1; -.
DR   PIR; S28052; S28052.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005989; Suc7H_symport.
DR   Pfam; PF00083; sugar_tr; 1.
DR   TIGRFAMs; TIGR01301; GPH_sucrose; 1.
KW   Transmembrane; Transport; Sugar transport; Symport.
FT   DOMAIN        1     37       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     38     58       1 (POTENTIAL).
FT   TRANSMEM     72     92       2 (POTENTIAL).
FT   TRANSMEM    107    127       3 (POTENTIAL).
FT   TRANSMEM    145    165       4 (POTENTIAL).
FT   TRANSMEM    184    204       5 (POTENTIAL).
FT   TRANSMEM    230    250       6 (POTENTIAL).
FT   TRANSMEM    295    315       7 (POTENTIAL).
FT   TRANSMEM    338    358       8 (POTENTIAL).
FT   TRANSMEM    373    393       9 (POTENTIAL).
FT   TRANSMEM    422    442       10 (POTENTIAL).
FT   TRANSMEM    455    475       11 (POTENTIAL).
FT   TRANSMEM    488    508       12 (POTENTIAL).
FT   DOMAIN      509    525       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   525 AA;  54992 MW;  018347A4D2CC1CC6 CRC64;

  Query Match             11.3%;  Score 324.5;  DB 1;  Length 525;
  Best Local Similarity   23.2%;  Pred. No. 2.1e-16;
  Matches  129;  Conservative   91;  Mismatches  196;  Indels  141;  Gaps   20;

Qy         23 LLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74
              |    ||    :|||: :        : | :  :|:   :   :    ||: |::  ||:|
Db         35 LKKLGLVASVAAGVQFGWALQLSLLTPYVQLLGIPHTWAAYIWLCGPISGMIVQPLVGYY 94

Qy         75 SDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPDP-----RPLELALLIL 127
              ||     |:|||||||  | :  : ::: ||    |: |   |      ||      :|  :|: ::
Db         95 SDRCTSRFGRRRPFIAAGAALVAVAVGLI---GFAADIGAASGDPTGNVAKPRAIAVFVV 151

Qy        128 GVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID---- 182
              |  :||          |  |||:|:         | | : ::| ::||    ||   :
Db        152 GFWILDVANNTLQGPCRALLADMAAGSQTKTRYANAFFSFFMALGNIGGYAAGSYSRLYT 211

Qy        183 ----WDTSALAPYLGTQEECLFGLLT-LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
                  |:|   |    : | |  :| || || :| :::   :  :  :  | |   :
Db        212 VPFPTKTAACDVYCANLKSCFFISITLLIVLTILALSVVKERQITIDEIQEEEDLKNRNN 271

Qy        238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
              | |    ||| |   | |:  |  |     |: :  | :    :|:|    | || ||::|
Db        272 SSGC----ARLPF--FGQLIGALKDL----PKPMLILLLVTALNWIAWFPFLLFDTDWMG 321

Qy        298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVY 355
              : :| |           : ||:||  |:||| :   :  | || :: | :  |    |
Db        322 KEVYGGT------VGEGKLYDQGVHAGALGLMINSVVLGVMSLSIEGLARMVGGAKRLWG 375
```

```
Qy        356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415
              : ::        | ||: :| :|| ||                   : |:    :
Db        376 IVNIIL------AVCLAMTV-LVTKSAE------------------HFRDSHHIM---- 405

Qy        416 GDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVG 475
                                                 || : ||||| |     ::: |:|
Db        406 ---------------------------------GSAVPPPPPA--GVKGGALAIFAVLG 429

Qy        476 EPTEARV-VP-----------GRGICLDLAILDSAFLLSQV----------------AP 506
              |      :|           | |  | | :|: | :: |:                  |
Db        430 IPLAITFSIPFALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPWDAMFGGGNLP 489

Qy        507 SLFMGSIVQLSQSVTAY 523
              :  :|::   : :| ::
Db        490 AFVVGAVAATASAVLSF 506
```

RESULT 6
PAST_HUMAN
ID PAST_HUMAN STANDARD; PRT; 509 AA.
AC Q9Y2W3;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Proton-associated sugar transporter A (PAST-A) (Deleted in
DE neuroblastoma 5 protein) (DNb-5) (Fragment).
GN DNB5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20195630; PubMed=10729226;
RA Amler L.C., Bauer A., Corvi R., Dihlmann S., Praml C., Cavenee W.K.,
RA Schwab M., Hampton G.M.;
RT "Identification and characterization of novel genes located at the
RT t(1;15)(p36.2;q24) translocation breakpoint in the neuroblastoma cell
RT line NGP.";
RL Genomics 64:195-202(2000).
CC -!- FUNCTION: Mediates glucose uptake along the pH gradient (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: Expressed in adult heart, brain, muscle and
CC kidney, with very strong expression in brain. Also expressed in
CC fetal brain, kidney and lung.
CC -!- SIMILARITY: Belongs to the glycoside-pentoside-hexuronide (GPH)
CC cation symporter transporter (TC 2.A.2) family.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AF118274; AAD27583.1; ALT_INIT.
DR Genew; HGNC:17939; DNB5.
DR MIM; 605763; -.
KW Transmembrane; Transport; Sugar transport; Symport.
FT NON_TER 1 1
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 294 314 POTENTIAL.
FT TRANSMEM 334 354 POTENTIAL.
FT TRANSMEM 361 381 POTENTIAL.
FT TRANSMEM 388 408 POTENTIAL.
FT TRANSMEM 446 466 POTENTIAL.
FT TRANSMEM 469 489 POTENTIAL.
SQ SEQUENCE 509 AA; 54637 MW; EEE766819D267C4C CRC64;

Query Match 9.3%; Score 267; DB 1; Length 509;
Best Local Similarity 25.3%; Pred. No. 3.2e-12;
Matches 72; Conservative 42; Mismatches 93; Indels 78; Gaps 4;

```
Qy        259 RLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYD 318
              ||    | ||: || | |      |::     ||||||:|| ::|| |:|   :|| : |:
Db        272 RLCSTICNMPKALRTLCVNHFLGWLSFEGMLLFYTDFMGEVVFQGDPKAPHTSEAYQKYN 331

Qy        319 EGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV 378
              || ||  |: :    :  :| ::::| :     | :|  :  || :  |   || :: ||
Db        332 SGVTMGCWGMCIYAFSAAFYSAILEKLEEFLSVRTLYFIAYLAFGLGTGLATLSRNLYVV 391

Qy        379 TASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG 438
              :    |  || |  |||:|   |:: |     |: |     || |
Db        392 LSLCITYGILFSTLCTLPYSLLCDYYQSK-----KFAG-----SSADGTR---------- 431

Qy        439 APFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSA 498
                                                          ||: :|:::|
Db        432 ------------------------------------------------RGMGVDISLLSCQ 444

Qy        499 FLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVF 543
              : |:|:  || :|           : || |     :||:: | |
Db        445 YFLAQILVSLVLGP----------LTSAVGSANGVMYFSSLVSF 478
```

RESULT 7
SUT1_SCHPO
ID SUT1_SCHPO STANDARD; PRT; 553 AA.
AC O14091; Q9UU60;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE General alpha-glucoside permease.
GN SUT1 OR SPAC2F3.08.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 35-166 FROM N.A.
RC STRAIN=968 h90;
RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,

```
RA   Hiraoka Y.;
RT   "Large-scale screening of intracellular protein localization in living
RT   fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL   Genes Cells 5:169-190(2000).
RN   [3]
RP   IDENTIFICATION, AND FUNCTION.
RX   MEDLINE=21065106; PubMed=11136464;
RA   Reinders A., Ward J.M.;
RT   "Functional characterization of the alpha-glucoside transporter Sut1p
RT   from Schizosaccharomyces pombe, the first fungal homologue of plant
RT   sucrose transporters.";
RL   Mol. Microbiol. 39:445-454(2001).
CC   -!- FUNCTION: Responsible for the transport of maltose and sucrose
CC       into the cell, with the concomitant export of a proton (Symport
CC       system).
CC   -!- PATHWAY: Maltose and sucrose metabolism.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Belongs to the glycoside-pentoside-hexuronide (GPH)
CC       cation symporter transporter (TC 2.A.2) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z99165; CAB16264.1; -.
DR   EMBL; AB027797; BAA87101.1; -.
DR   PIR; T38541; T38541.
DR   GeneDB_SPombe; SPAC2F3.08; -.
KW   Transmembrane; Transport; Sugar transport; Symport.
FT   DOMAIN        1     33       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     34     54       POTENTIAL.
FT   DOMAIN       55     72       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     73     93       POTENTIAL.
FT   DOMAIN       94    111       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    112    132       POTENTIAL.
FT   DOMAIN      133    140       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    141    161       POTENTIAL.
FT   DOMAIN      162    186       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    187    207       POTENTIAL.
FT   DOMAIN      208    216       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    217    237       POTENTIAL.
FT   DOMAIN      238    280       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    281    301       POTENTIAL.
FT   DOMAIN      302    322       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    323    343       POTENTIAL.
FT   DOMAIN      344    424       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    425    445       POTENTIAL.
FT   DOMAIN      446    452       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    453    473       POTENTIAL.
FT   DOMAIN      474    494       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    495    515       POTENTIAL.
FT   DOMAIN      516    521       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    522    542       POTENTIAL.
FT   DOMAIN      543    553       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   553 AA;  61744 MW;  2C31DFC2D88A7C7A CRC64;

  Query Match             8.2%;  Score 235.5;  DB 1;  Length 553;
  Best Local Similarity   24.0%;  Pred. No. 6.8e-10;
  Matches   82;  Conservative   59;  Mismatches  153;  Indels   47;  Gaps   10;

Qy         19 LLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDHW 78
              | |:||  |::: : : | | | :|: ::: :::   ||: |::  |: | ||
Db         39 LTVSLL--GVQLTWSVELGYGTPYLFSLGLRKEWTSIIWIAGPLTGILIQPIAGILSDRV 96

Qy         79 RGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPD-----------PRPLELALLIL 127
                 | ||||||:   ||    ||||:   ||      ||           : : : |  :
Db         97 NSRIGRRRPFMLCASLLGTFSLFLM---GW-----APDICLFIFSNEVLMKRVTIVLATI 148

Qy        128 GVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAID-WDTS 186
               : |||    |      :|: |  |  |   :| |    || :|  ||||  :  :
Db        149 SIYLLDVAVNVVMASTRSLIVDSVRS-DQQHEANSWAGRMIGVGNVLGYLLGYLPLYRIF 207

Qy        187 ALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRA 246
              :   :   | |:   ::|: ||    |: |:|     | |  : ::
Db        208 SFLNFTQLQVFCVLASISLV-LTVTITTIFVSER-RFPPVEHEKSVAG------------ 253

Qy        247 RLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPR 306
                    :   |    :| ||:|:   :  ::      |  : | :||    :  |:
Db        254 -----EIFEFFTTMRQSITALPFTLKRICFVQFFAYFGWFPFLFYITTYVGILYLRHAPK 308

Qy        307 AEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQ 347
                    :|   | ||  | |   ||| :  :  |::
Db        309 GH-----EEDWDMATRQGSFALLLFAIISLAANTALPLLLE 344


RESULT 8
YD74_SYNY3
ID   YD74_SYNY3     STANDARD;      PRT;   544 AA.
AC   P74168;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical symporter sll1374.
GN   SLL1374.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY
CC       (SGF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90912; BAA18257.1; -.
DR   PIR; S75696; S75696.
DR   InterPro; IPR001927; Na/Gal_symport.
DR   TIGRFAMs; TIGR00792; gph; 1.
DR   PROSITE; PS00872; NA_GALACTOSIDE_SYMP; FALSE_NEG.
KW   Hypothetical protein; Transport; Transmembrane; Symport;
KW   Complete proteome.
FT   TRANSMEM     31     51       POTENTIAL.
FT   TRANSMEM     52     72       POTENTIAL.
FT   TRANSMEM     84    104       POTENTIAL.
FT   TRANSMEM    116    136       POTENTIAL.
FT   TRANSMEM    162    182       POTENTIAL.
FT   TRANSMEM    191    211       POTENTIAL.
FT   TRANSMEM    230    250       POTENTIAL.
FT   TRANSMEM    257    277       POTENTIAL.
FT   TRANSMEM    318    338       POTENTIAL.
FT   TRANSMEM    356    376       POTENTIAL.
FT   TRANSMEM    383    403       POTENTIAL.
FT   TRANSMEM    407    427       POTENTIAL.
FT   TRANSMEM    450    470       POTENTIAL.
FT   TRANSMEM    501    521       POTENTIAL.
```

```
SQ   SEQUENCE   544 AA;  60064 MW;  C91D0EDFF32277EE CRC64;

  Query Match             5.0%;  Score 142;  DB 1;  Length 544;
  Best Local Similarity   22.3%;  Pred. No. 0.0045;
  Matches  105;  Conservative   60;  Mismatches  179;  Indels  126;  Gaps   20;

Qy         16 AQLLLVNLLTFGLEVCLAAG-----------ITYVPPLLLEV-GVEEKFMTMVLGIGPVL 63
Db          2 SQSLSAEKLHFTTKLAYGAGDFGPAITANILVFYLLFFLTDVAGIPAALAGSVLMIGKIF 61

Qy         64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELA 123
Db         62 DAINDPIIGLLSDRTRSRWGRRLPWM----LGGMIPPALFYTAQWLIPHFSDDRLTNQWG 117

Qy        124 LLILGVGLLDFCGQVCFT----PLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLP 179
Db        118 LFIYYVAIA-MAFNLCYTTVNLPYTALTPELTQNYNE-RTRLNSFRFAFSIGGSILSLIL 175

Qy        180 AIDWDTSALAPYLGTQEECLFGLL-TLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
Db        176 YI-----LIAAGLPDRPQQQFGELGVMISVLSISALLWSALRLQEKGKEP---ILSPSL- 226

Qy        239 PHCCPCRARLA------------------FRNLGAL------------------------ 256
Db        227 ------RRRLAPLLMAAGITLILLAIAKSFNLLGGSGFDYISFFLILLGLIWGGFGFTLR 280

Qy        257 ----------------------LPRLHQLCCRMPRTLRR-LFV--AELCSWMALMTFTLF 291
Db        281 DSAVEEHLQKLENSPSPGVTENLPLLKQL--KIAFSNRAFLFVIGIYLCSWLAVQLTASI 338

Qy        292 YTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGT 351
Db        339 LVYFVVS--WMGLNEQQSGT--------------IALAVQ-GTALVMLFVWQALAQFLDK 381

Qy        352 RAVY-LASVAAFPVAAGATCLS-HSVAVVTASAALTGFTFSALQILPYTL 399
Db        382 KVIYFLGSMVWMGAEAGLWLVQPGQVALLYTLAIFAGVGVSVAYLIPWSM 431


RESULT 9
GT10_HUMAN
ID   GT10_HUMAN     STANDARD;      PRT;   541 AA.
AC   O95528; Q9H4I6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 10
DE   (Glucose transporter type 10).
GN   SLC2A10 OR GLUT10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=21145593; PubMed=11247674;
RA   McVie-Wylie A.J., Lamson D.R., Chen Y.T.;
RT   "Molecular cloning of a novel member of the GLUT family of
RT   transporters, SLC2A10 (GLUT10), localized on chromosome 20q13.1: a
RT   candidate gene for NIDDM susceptibility.";
RL   Genomics 72:113-117(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Stavrides G.S., Hashim Y., Huckle E.J., Deloukas P.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Fossey S.C., Mihic S.J., Craddock A.L., Mychaleckyj J.C., Dawson P.A.,
RA   Bowden D.W.;
RT   "GLUT10: a novel glucose transporter in the type 2 diabetes linked
RT   region of chromosome 20q12-13.1.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21638749; PubMed=11780052;
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
CC   -!- FUNCTION: Facilitative glucose transporter (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Widely expressed; highest levels in liver and
CC       pancreas.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC       TRANSPORTERS SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF321240; AAK26294.1; -.
DR   EMBL; AL137188; CAB69822.2; -.
DR   EMBL; AF248053; AAK31911.1; -.
DR   EMBL; AL031055; CAA19926.2; -.
DR   Genew; HGNC:13444; SLC2A10.
DR   MIM; 606145; -.
DR   GO; GO:0016021; C:integral to membrane; NAS.
DR   GO; GO:0005351; F:sugar porter activity; NAS.
DR   GO; GO:0015758; P:glucose transport; NAS.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR003663; Sugar_transpt.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW   Transport; Sugar transport; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN        1      15       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      16     36       1 (POTENTIAL).
FT   DOMAIN        37     48       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      49     69       2 (POTENTIAL).
FT   DOMAIN        70     77       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      78     98       3 (POTENTIAL).
FT   DOMAIN        99     106      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    107     127       4 (POTENTIAL).
```

```
FT   DOMAIN      128    134       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    135    155       5 (POTENTIAL).
FT   DOMAIN      156    166       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    167    187       6 (POTENTIAL).
FT   DOMAIN      188    233       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    234    254       7 (POTENTIAL).
FT   DOMAIN      255    269       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    270    290       8 (POTENTIAL).
FT   DOMAIN      291    298       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    299    319       9 (POTENTIAL).
FT   DOMAIN      320    414       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    415    435       10 (POTENTIAL).
FT   DOMAIN      436    445       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    446    466       11 (POTENTIAL).
FT   DOMAIN      467    476       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    477    497       12 (POTENTIAL).
FT   DOMAIN      498    541       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    334    334       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   541 AA;  56911 MW;  6D644525FA136908 CRC64;

  Query Match             4.8%;  Score 136;  DB 1;  Length 541;
  Best Local Similarity   21.9%;  Pred. No. 0.012;
  Matches  130;  Conservative   72;  Mismatches  210;  Indels  182;  Gaps   27;

Qy         16 AQLLLVNLLTFGLEVCLAAGITYVPPLLLEVG---VEEKFMTMVLGIGPVLGLVCVPLLG 72
              | : |:  |||| |: : :|   : || |: |   :|::|:    | :| :|  :     |
Db         12 ASVSLLGGLTFGYELAVISGA--LLPLQLDFGLSCLEQEFLVGSLLLGALLASLVGGFLI 69

Qy         73 SASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVGLL 132
                    |||::  : : :| :|     :  ||              || |:||  ::
Db         70 DC-------YGRKQAILGS-NLVLLAGSLTLGLAG------------SLAWLVLGRAVV 108

Qy        133 DFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLG----GCLGYLLPAIDWDTS-- 186
               |   :         :|:|   |        |:|   |::|       | | |    |
Db        109 GFAISLSSMACCIYVSELV-GPRQRGVLVSLYEAGITVGILLSYALNYALAGTPWGWRHM 167

Qy        187 ---ALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCP 243
                 | ||         :   |:|:||   | |     | |       | :|  || | |
Db        168 FGWATAP-------AVLQSLSLLFLP--AGT---DETATHKDLIPLQGGEAPKLGPG--- 212

Qy        244 CRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQG 303
               | | :| :|      :   |  | : :|     :: :     | |   | : |
Db        213 -RPRYSFLDLFRARDNMRG---RTTVGLGLVLFQQLTGQPNVLCYA--STIFSSVGFHGG 266

Qy        304 VPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA--- 360
                          : |:||    |: :  :|    || | | ||: ||  |
Db        267 --------------SSAVLASVGL---GAVKVAATLTAMGLVDRAGRRALLLAGCALMA 308

Qy        361 -----------AFPVAAGATCLSHSVAVVTASAALTG----FTFSALQILPYTLASLYHR 405
                        | |: :| :||  :|    |     | |       |:|  :| |
Db        309 LSVSGIGLVSFAVPMDSGPSCL--AVPNATGQTGLPGDSGLLQDSSLPPIPRT------N 360

Qy        406 EKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASA 465
               | |             |  | |:     | | :  |:         |  || ||
Db        361 EDQ--------------REPILSTAKKTKPHPRSGDPSAPPRLALSSALPGPP------- 399

Qy        466 CDVSVRVVVGEPTEARVVPGRG-------ICLDLAILDSAF----------LLSQVAP-- 506
                          :| ||       | | :  |||           :||:: |
Db        400 ----------------LPARGHALLRWTALLCLMVFVSAFSFGFGPVTWLVLSEIYPVE 442

Qy        507 ---------------SLFMG-------SIVQLSQSVTAYMVSAA-GLGLVAIY 536
                            :||:          : || :   | ::|  ||| : ::
Db        443 IRGRAFAFCNSFNWAANLFISLSFLDLIGTIGLSWTFLLYGLTAVLGLGFIYLF 496


RESULT 10
TCR1_ECOLI
ID   TCR1_ECOLI     STANDARD;      PRT;   399 AA.
AC   P02982;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Tetracycline resistance protein, class A (TETA(A)).
GN   TETA.
OS   Escherichia coli.
OG   Plasmid RP1.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=Tn1721;
RX   MEDLINE=92192465; PubMed=1312499;
RA   Allmeier H., Cresnar B., Greck M., Schmitt R.;
RT   "Complete nucleotide sequence of Tn1721: gene organization and a
RT   novel gene product with features of a chemotaxis protein.";
RL   Gene 111:11-20(1992).
RN   [2]
RP   SEQUENCE OF 85-399 FROM N.A.
RC   STRAIN=DH1;
RA   Fletcher J.N., Hart C.A., Batt R.M., Saunders J.R.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   PLASMID=RP1; TRANSPOSON=Tn1721;
RX   MEDLINE=83299270; PubMed=6310527;
RA   Waters S.H., Rogowsky P., Grinsted J., Altenbuchner J., Schmitt R.;
RT   "The tetracycline resistance determinants of RP1 and Tn1721:
RT   nucleotide sequence analysis.";
RL   Nucleic Acids Res. 11:6089-6105(1983).
RN   [4]
RP   TOPOLOGY.
RX   MEDLINE=92388137; PubMed=1517220;
RA   Allard J.D., Bertrand K.P.;
RT   "Membrane topology of the pBR322 tetracycline resistance protein.
RT   TetA-PhoA gene fusions and implications for the mechanism of TetA
RT   membrane insertion.";
RL   J. Biol. Chem. 267:17809-17819(1992).
CC   -!- FUNCTION: Resistance to tetracycline by an active tetracycline
CC       efflux. This is an energy-dependent process that decreases the
CC       accumulation of the antibiotic in whole cells. This protein
CC       functions as a metal-tetracycline/H(+) antiporter.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X61367; CAA43643.1; -.
DR   EMBL; L29404; AAA83545.1; -.
DR   EMBL; X00006; CAA24909.1; -.
DR   PIR; A03509; YTECR1.
DR   PIR; JQ1479; JQ1479.
DR   InterPro; IPR004734; Drug_resist.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   InterPro; IPR001958; TCR_TetA.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR01035; TCRTETA.
DR   TIGRFAMs; TIGR00880; 2_A_01_02; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW   Antibiotic resistance; Transmembrane; Inner membrane; Transport;
KW   Antiport; Ion transport; Hydrogen ion transport; Transposable element;
KW   Plasmid.
FT   DOMAIN        1      7       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM      8     27       PROBABLE.
FT   DOMAIN       28     45       PERIPLASMIC (PROBABLE).
FT   TRANSMEM     46     66       PROBABLE.
```

```
FT   DOMAIN       67     79       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM     80    100       PROBABLE.
FT   DOMAIN      101    103       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    104    124       PROBABLE.
FT   DOMAIN      125    138       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    139    159       PROBABLE.
FT   DOMAIN      160    160       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    161    181       PROBABLE.
FT   DOMAIN      182    210       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    211    231       PROBABLE.
FT   DOMAIN      232    246       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    247    267       PROBABLE.
FT   DOMAIN      268    277       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    278    298       PROBABLE.
FT   DOMAIN      299    299       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    300    320       PROBABLE.
FT   DOMAIN      321    339       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    340    360       PROBABLE.
FT   DOMAIN      361    364       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    365    385       PROBABLE.
FT   DOMAIN      386    399       CYTOPLASMIC (PROBABLE).
FT   CONFLICT      5      5       R -> I (IN REF. 3).
FT   CONFLICT     55     55       M -> V (IN REF. 3).
FT   CONFLICT     75     75       V -> I (IN REF. 3).
FT   CONFLICT     84     84       A -> T (IN REF. 3).
FT   CONFLICT    201    203       ASF -> SFV (IN REF. 3).
SQ   SEQUENCE   399 AA;  42240 MW;  298427E6B5478374 CRC64;

  Query Match             4.4%;  Score 125.5;  DB 1;  Length 399;
  Best Local Similarity   22.3%;  Pred. No. 0.052;
  Matches  104;  Conservative   57;  Mismatches  136;  Indels  169;  Gaps   27;

Qy         19 LLVNLLTFGLEVCLAAGITYVPPL---LLEVGVEEKFMT----MVLGIGPVLGLVCVPLL 71
              |:| | |  |:   | ||  : |:   ||   |    :|    ::| :  ::   | |:|
Db          7 LIVILSTVALD---AVGIGLIMPVLPGLLRDLVHSNDVTAHYGILLALYALMQFACAPVL 63

Qy         72 GSASDHWRGRYGRR------------------RPFIWALSLGILLSLFLIPRAGWLAGL 112
              |: ||    |:|||                  ||:| | :| :::
Db         64 GALSD----RFGRRPVLLVSLAGAAVDYAIMATAPFLWVLYIGRIVA------------ 106

Qy        113 LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGG 172
                          |:    | |      | ::|:    :  |     : ||
Db        107 ----------------GITGATGAVA----GAYIADITDGDERARH----FGFM---SA 138

Qy        173 CLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGL 232
              | |: : |          | ||        ||:
Db        139 CFGFGMVA--------GPVLG-------GLM---------------------------- 154

Qy        233 SAPSLSPHCCPCRARLAFRNLG-----ALLPRLHQLCCRMPRTLRRLFVAELCS--W--- 282
                |||  |  |  |   |       |||  |:      | |||  :  | |  |
Db        155 --GGFSPH-APFFAAAALNGLNFLTGCFLLPESHK---GERRPLRREALNPLASFRWARG 208

Qy        283 ---MALMTFTLFYTDFVGEGLYQGVPRA--EPGTEARRHYDE---GVRMGSLGLFLQCAI 334
                :| :    |   ||:     || |     | | |:|    |: : : |:    |
Db        209 MTVVAALMAVFFIMQLVGQ-----VPAALWVIFGEDRFHWDATTIGISLAAFGILHSLAQ 263

Qy        335 SLVFSLVMDRLVQRFGTRAVYLASVA---AFPVAAGAT--CLSHSVAVVTASAALTGFTF 389
              :::   |  || :|   ||: |  :|    : : | ||   ::  : |: ||    |
Db        264 AMITGPVAARLGER---RALMLGMIADGTGYILLAFATRGWMAFPIMVLLASG---GIGM 317

Qy        390 SALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGP 435
               ||| :   |:     |:|       |   |: :  : :|| : ||
Db        318 PALQAM---LSRQVDEERQ-------GQLQGSLAALTSLTSIV-GP 352


RESULT 11
PUR8_STRLP
ID   PUR8_STRLP     STANDARD;      PRT;   503 AA.
AC   P42670;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Puromycin resistance protein pur8.
GN   PUR8.
OS   Streptomyces lipmanii (Streptomyces alboniger).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=132472;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12461;
RX   MEDLINE=94109397; PubMed=7916693;
RA   Tercero J.A., Lacalle R.A., Jimenez A.;
RT   "The pur8 gene from the pur cluster of Streptomyces alboniger encodes
RT   a highly hydrophobic polypeptide which confers resistance to
RT   puromycin.";
RL   Eur. J. Biochem. 218:963-971(1993).
CC   -!- FUNCTION: MAY BE INVOLVED IN ACTIVE PUROMYCIN EFFLUX ENERGIZED BY
CC       A PROTON-DEPENDENT ELECTROCHEMICAL GRADIENT. IN ADDITION, IT COULD
CC       BE IMPLICATED IN SECRETING N-ACETYLPUROMYCIN, THE LAST
CC       INTERMEDIATE OF THE PUROMYCIN BIOSYNTHESIS PATHWAY, TO THE
CC       ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X76855; CAA54186.1; -.
DR   InterPro; IPR004638; Efflux_EmrB.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR001411; TCR_TetB.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR01036; TCRTETB.
DR   TIGRFAMs; TIGR00711; efflux_EmrB; 1.
DR   PROSITE; PS50850; MFS; 1.
KW   Antibiotic resistance; Antibiotic biosynthesis; Transmembrane;
KW   Transport.
FT   DOMAIN       1     24       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    25     45       POTENTIAL.
FT   DOMAIN      46     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    65     85       POTENTIAL.
FT   DOMAIN      86     92       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    93    113       POTENTIAL.
FT   DOMAIN     114    122       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   123    143       POTENTIAL.
FT   DOMAIN     144    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   153    173       POTENTIAL.
FT   DOMAIN     174    181       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   182    202       POTENTIAL.
FT   DOMAIN     203    212       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   213    233       POTENTIAL.
FT   DOMAIN     234    241       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   242    262       POTENTIAL.
FT   DOMAIN     263    280       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   281    301       POTENTIAL.
FT   DOMAIN     302    313       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   314    334       POTENTIAL.
FT   DOMAIN     335    346       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   347    367       POTENTIAL.
FT   DOMAIN     368    371       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   372    392       POTENTIAL.
FT   DOMAIN     393    422       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   423    443       POTENTIAL.
FT   DOMAIN     444    461       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   462    482       POTENTIAL.
FT   DOMAIN     483    503       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   503 AA;  51852 MW;  FA3B814DF9A3EB74 CRC64;
```

```
  Query Match             4.2%;  Score 121.5;  DB 1;  Length 503;
  Best Local Similarity   23.2%;  Pred. No. 0.13;
  Matches  131;  Conservative   61;  Mismatches  175;  Indels  197;  Gaps   32;

Qy         26 FGLEVCLAAGITYV----------PPLLLEVGVEEKFMTMV-----LGIGPVLGLVCVPL 70
              :|| | ||| :  |           | :  ::|: :      |      |  | :|      |
Db         26 WGLVVILAAQLLVVLDGTVVNIALPSVQRDLGMSDTSRQWVITAYTLAFGGLL------L 79

Qy         71 LGSASDHWRGR----YGRRRPFIWALSLGILLSLFLIPRAGW-LAGLL---CPDPRPLEL 122
              ||        ||    :|||| |    ::|||          |: || ||    |||  | |
Db         80 LG-------GRVADAFGRRRIF----AVGIL---------GFGLASLLGGAAPDPGTLFL 119

Qy        123 ALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAID 182
              |  : ||    |   :    | ||:: || :|    :|: ||  :   |  :| |
Db        120 ARALQGV----FAAALAPAAL-ALINTLFTEPGERGKAFGVYGAVSGGGAAVGLLAG--- 171

Qy        183 WDTSALAPYLGTQEECLFGLLT--LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
                              ||||  | :  |:      ||  | ||
Db        172 ------------------GLLTEYLDWRWCLYVNAPVALLALLG---------------- 197

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAEL--CSWMALMTFTLFYTDFVGE 298
                 ||           ||||       |  | :|       |  |  :  : :
Db        198 ---CR----------LLPRDR----RTGRAVRLDLPGTLLGCGGLVAIVYAF-------- 232

Qy        299 GLYQGVPRAEPGTEARRHYDEG--VRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYL 356
                           ||    : :   ||:  ||: :  | :||    | | |:
Db        233 -----------AEAESGWGDPLVVRLLVLGVLMLVAFALVERRVQDPLLP--------- 271

Qy        357 ASVAAFPVAAGATCLSHSVAVVTASAALTG----FTFSALQILPYT--LASLYHREKQVF 410
                | |  |  |:      : |      | |       |:      || |:  |  :      |
Db        272 PGVVAHRVRGGS-----FLVVGLPQIGLFGLFLFLTYYLQGILDYSPVLTGV------AF 320

Qy        411 LPKYRGDTGGASSEDSLMTSFLPGPKP-----GAPFPNGHVGAGGSGLL----PPPPALC 461
              ||   |   |:|    :    ||  :|     ||          | |  ||     |  |
Db        321 LPLGLGIAVGSS---LIAARLLPRTRPRTLIVGALL----AAAAGMALLTRLEPDTP--- 370

Qy        462 GASACDVSVRVVVGEPTEARVVPGRGI-CLDLAILDSAFLLSQVAP------SLFMGSIV 514
                        :| :     |::: | || |: :  : :|   ::|||      :  : |
Db        371 ---------QVYLTHLLPAQILIGLGIGCMMMPAMHTA--TARVAPHEAGAAAAVVNSAQ 419

Qy        515 QLSQSVTAYMVSAAGLGLVAIYFA 538
              |:  ::   :::    |  | | |
Db        420 QVGGALGVALLNTVSTGATAAYLA 443


RESULT 12
AMPG_ECOLI
ID   AMPG_ECOLI     STANDARD;      PRT;   491 AA.
AC   P36670;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   AmpG protein.
GN   AMPG OR B0433 OR Z0536 OR ECS0487.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94049112; PubMed=8231804;
RA   Lindquist S., Weston-Hafer K., Schmidt H., Pul C., Korfmann G.,
RA   Erickson J., Sanders C., Martin H.H., Normark S.;
RT   "AmpG, a signal transducer in chromosomal beta-lactamase induction.";
RL   Mol. Microbiol. 9:703-715(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SN0301-1, SN0301-3, and SN0301-5;
RX   MEDLINE=95291453; PubMed=7773404;
RA   Schmidt H., Korfmann G., Barth H., Martin H.H.;
RT   "The signal transducer encoded by ampG is essential for induction of
RT   chromosomal AmpC beta-lactamase in Escherichia coli by beta-lactam
RT   antibiotics and 'unspecific' inducers.";
RL   Microbiology 141:1085-1092(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA   Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: Probably acts as a permease in the beta-lactamase
CC       induction system and in peptidoglycan recycling.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- SIMILARITY: TO H.INFLUENZAE HI0350 AND TO YEAST YBR220C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S67816; AAB28884.1; -.
DR   EMBL; X82158; CAA57651.1; -.
DR   EMBL; X82159; CAA57652.1; -.
DR   EMBL; X82160; CAA57653.1; -.
DR   EMBL; AE000149; AAC73536.1; -.
DR   EMBL; U82664; AAB40189.1; -.
DR   EMBL; AE005222; AAG54783.1; -.
DR   EMBL; AP002551; BAB33910.1; -.
DR   PIR; C85540; C85540.
DR   PIR; G90689; G90689.
DR   PIR; S37391; S37391.
DR   EcoGene; EG12183; ampG.
DR   InterPro; IPR004752; AmpG_permease.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   TIGRFAMs; TIGR00901; 2A0125; 1.
DR   PROSITE; PS50850; MFS; 1.
```

```
KW   Transport; Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM     13     33       POTENTIAL.
FT   TRANSMEM     81    105       POTENTIAL.
FT   TRANSMEM    110    128       POTENTIAL.
FT   TRANSMEM    174    197       POTENTIAL.
FT   TRANSMEM    224    244       POTENTIAL.
FT   TRANSMEM    266    283       POTENTIAL.
FT   TRANSMEM    328    345       POTENTIAL.
FT   TRANSMEM    382    406       POTENTIAL.
FT   TRANSMEM    425    450       POTENTIAL.
FT   TRANSMEM    461    485       POTENTIAL.
FT   VARIANT     151    151       G -> D (IN NON-INDUCIBLE MUTANT
FT                                SN0301-1).
FT   VARIANT     268    268       G -> D (IN NON-INDUCIBLE MUTANT
FT                                SN0301-3).
FT   VARIANT     373    373       G -> D (IN NON-INDUCIBLE MUTANT
FT                                SN0301-5).
SQ   SEQUENCE   491 AA;  53245 MW;  6B0D7E2BC699CA9C CRC64;

  Query Match             4.2%;  Score 119.5;  DB 1;  Length 491;
  Best Local Similarity   21.0%;  Pred. No. 0.18;
  Matches  126;  Conservative   76;  Mismatches  193;  Indels  205;  Gaps   30;

Qy          9 RLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTM------VLGIGPV 62
              |: :  :: :||:        :  |:|:    || |  |  : :||:       :|   :
Db          7 RIFQQPRSAILLI--------LGFASGL----PLALTSGTLQAWMTVENIDLKTIGFFSL 54

Qy         63 LGLVCV------PLLGSASDHWRGRYGRRRPFIWALSLGILLSL----FLIP--RAGWLA 110
              :|   |      ||:   :  :   :|||| :: |  : :|:::    || |  :  |:|
Db         55 VGQAYVFKFLWSPLMDRYTPPF---FGRRRGWLLATQILLLVAIAAMGFLEPGTQLRWMA 111

Qy        111 GLLCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDH-CRQAYSVYAF--- 166
              |            :: ||        :|  :|:    :      | ||  :
Db        112 ALAV----------------VIAFCSASQDIVFDAWKTDVLPAEERGAGAAISVLGYRLG 155

Qy        167 MISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPT 226
              |:  ||   :|  | |        || |   :: |:  : : |: |||| |     ||
Db        156 MLVSGGLALWL--ADKW--------LGWQG--MYWLMAALLIPCIIATLLAPE-----PT 198

Qy        227 EPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAEL------- 279
              :                          :|        :|:|| :  || |
Db        199 D--------------------------TIP--------VPKTLEQAVVAPLRDFFGRN 222

Qy        280 CSWMALMTFTL------FYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCA 333
              :|: |:   |      |        | :||     | :|    : ||   :|||
Db        223 NAWLILLLIVLYKLGDAFAMSLTTTFLIRGV-----GFDAG---EVGVVNKTLGLLATIV 274

Qy        334 ISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQ 393
              :|   ::| ||      ||: :  :                    | | : :
Db        275 GALYGGILMQRLSL---FRALLIFGI----------------------LQGASNAGYW 307

Qy        394 ILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGL 453
              :|  |   ||     ||     |  | ::    |||              |     |
Db        308 LLSITDKHLYSMGAAVFFENLCGGMGTSAFVALLMT----------LCNKSFSATQFAL 356

Qy        454 LPPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVA---PSLFM 510
              |         ||     || || |    |   |          | | |  ||    | | :
Db        357 L---------SALSAVGRVYVG-PVAGWFVEAHGW--------STFYLFSVAAAVPGLIL 398

Qy        511 GSIVQLSQ----------SVTAYMVS--------AAGLGLVAIYFATQVVFDKSDLAKYS 552
                : : :           | |||           |||: |:|::    :  |  ||  :|
Db        399 LLVCRQTLEYTRVNDNFISRTAYPAGYAFAMWTLAAGVSLLAVWLLL-LTMDALDLTHFS 457


RESULT 13
YHJE_ECOLI
ID   YHJE_ECOLI     STANDARD;      PRT;   440 AA.
AC   P37643;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical metabolite transport protein yhjE.
GN   YHJE OR B3523.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94316500; PubMed=8041620;
RA   Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT   "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT   region from 76.0 to 81.5 minutes.";
RL   Nucleic Acids Res. 22:2576-2586(1994).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. STRONG, TO
CC       H.INFLUENZAE HI0281 AND HI0418.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00039; AAB18499.1; -.
DR   EMBL; AE000429; AAC76548.1; -.
DR   PIR; S47743; S47743.
DR   EcoGene; EG12249; yhjE.
DR   InterPro; IPR004736; Cit_H_symport.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005829; Sug_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   TIGRFAMs; TIGR00883; 2A0106; 1.
DR   PROSITE; PS50850; MFS; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW   Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   DOMAIN       1     34       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    35     55       1 (POTENTIAL).
FT   DOMAIN      56     66       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    67     87       2 (POTENTIAL).
FT   DOMAIN      88    108       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   109    129       3 (POTENTIAL).
FT   TRANSMEM   130    150       4 (POTENTIAL).
FT   DOMAIN     151    167       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   168    188       5 (POTENTIAL).
FT   DOMAIN     189    192       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM   193    213       6 (POTENTIAL).
FT   DOMAIN     214    248       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   249    269       7 (POTENTIAL).
FT   DOMAIN     270    289       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM   290    310       8 (POTENTIAL).
FT   DOMAIN     311    320       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   321    341       9 (POTENTIAL).
FT   DOMAIN     342    345       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM   346    366       10 (POTENTIAL).
FT   DOMAIN     367    384       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   385    405       11 (POTENTIAL).
FT   DOMAIN     406    410       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM   411    431       12 (POTENTIAL).
FT   DOMAIN     432    440       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   440 AA;  47208 MW;  3A23302A9DCBE5BE CRC64;

  Query Match             4.1%;  Score 117;  DB 1;  Length 440;
  Best Local Similarity   28.4%;  Pred. No. 0.24;
  Matches   48;  Conservative   29;  Mismatches   54;  Indels   38;  Gaps   10;

Qy         53 MTMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLI-PRAGWLAG 111
                |  |:| |     | ||: |  :|     :|||:  :   :| || :||   | |
Db        295 MMAVIGFG-----VMVPVAGLLAD----AFGRRKSMVIITTLIILFALFAFNPLLG---- 341
```

```
Qy        112 LLCPDPRP-LELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISL 170
              | |  | |:|: |:     : | |: ||| :||  |   |  |:  :|  ::
Db        342 ----SGNPILVFAFLLLGLSLMG----LTFGPMGALLPELF--PTEVR--YTGASFSYNV 389

Qy        171 GGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTL-IFLTCVAATLLVA 218
                ||          :::|||:  . :  :||  : ::|  :|   |:|
Db        390 ASILG----------ASVAPYIAAWLQTNYGLGAVGLYLAAMAGLTLIA 428


RESULT 14
SMVA_SALTY
ID   SMVA_SALTY     STANDARD;      PRT;   495 AA.
AC   P37594;
DT   01-OCT-1994 (Rel. 30, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Methyl viologen resistance protein smvA.
GN   SMVA OR STM1574.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SL1303;
RX   MEDLINE=95011654; PubMed=7926834;
RA   Hongo E., Morimyo M., Mita K., Machida I., Hama-Inaba H., Tsuji H.,
RA   Ichimura S., Noda Y.;
RT   "The methyl viologen-resistance-encoding gene smvA of Salmonella
RT   typhimurium.";
RL   Gene 148:173-174(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D26057; BAA05055.1; -.
DR   EMBL; AE008769; AAL20492.1; -.
DR   StyGene; SG10384; smvA.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR001411; TCR_TetB.
DR   PRINTS; PR01036; TCRTETB.
DR   PROSITE; PS50850; MFS; 1.
KW   Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM      5     25       POTENTIAL.
FT   TRANSMEM     44     64       POTENTIAL.
FT   TRANSMEM     73     93       POTENTIAL.
FT   TRANSMEM     96    116       POTENTIAL.
FT   TRANSMEM    135    155       POTENTIAL.
FT   TRANSMEM    158    178       POTENTIAL.
FT   TRANSMEM    192    212       POTENTIAL.
FT   TRANSMEM    220    240       POTENTIAL.
FT   TRANSMEM    260    280       POTENTIAL.
FT   TRANSMEM    299    319       POTENTIAL.
FT   TRANSMEM    327    347       POTENTIAL.
FT   TRANSMEM    357    377       POTENTIAL.
FT   TRANSMEM    391    411       POTENTIAL.
FT   TRANSMEM    469    489       POTENTIAL.
FT   CONFLICT    144    144       A -> R (IN REF. 1).
FT   CONFLICT    147    147       G -> V (IN REF. 1).
FT   CONFLICT    182    182       V -> D (IN REF. 1).
FT   CONFLICT    198    198       A -> V (IN REF. 1).
FT   CONFLICT    227    227       F -> Y (IN REF. 1).
FT   CONFLICT    447    447       L -> Y (IN REF. 1).
FT   CONFLICT    454    457       AILD -> GKYLT (IN REF. 1).
SQ   SEQUENCE   495 AA;  52134 MW;  961F77C748CDA164 CRC64;

  Query Match             4.1%;  Score 117;  DB 1;  Length 495;
  Best Local Similarity   21.0%;  Pred. No. 0.27;
  Matches   92;  Conservative   60;  Mismatches  151;  Indels  136;  Gaps   22;

Qy         17 QLLLVNLLTFGL-----------------EVCLAAGITYVPPLLLEVGV-----EEKFMT 54
              :||::   |||                  | || |   : | |  |:     |||
Db         72 RLLMLGGTLFGLASLAAAFSHTASWLIATRVLLAIGAAMIVPATL-AGIRATFCEEKHRN 130

Qy         55 MVLGI-----------GPVLG-------------LVCVPL----LGSASDHWRGRYGRR- 85
              | ||:           ||::|            |: ||:    :|  : :   : |||
Db        131 MALGVWAAVGSGGAAFGPLIGGILLEHFYWGSVFLINVPIVLVVMGLTARYVPRQAGRRD 190

Qy         86 RPF--------------------------IWALSL-----GILLSLFLIPRAGWLAGL 112
              :|                         :| :|       :|| ||:  :      :
Db        191 QPLNLGHAVMLIIAILLLVYSAKTALKGHLSLWVISFTLLTGALLLGLFIRTQLATSRPM 250

Qy        113 LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMI---- 168
              :  | |     ::: || ::    :     | |::   :   |    |    ||:
Db        251 I--DMRLFTHRIILSGV-VMAMTAMITLVGFELLMAQELQFV-HGLSPYEAGVFMLPVMV 306

Qy        169 ------SLGGCL------------GYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTC 210
                   : | |            |  | |: :   |:  :  ||:   :||: |:  :
Db        307 ASGFSGPIAGVLVSRLGLRLVATGGMALSALSFYGLAMTDF-STQQWQAWGLMALLGFS- 364

Qy        211 VAATLLVAEEAALGPTEPAE-GLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPR 269
              ||: |:|  :|:    |||   :| ::        | |     | ||| |      :|
Db        365 -AASALLASTSAIMAAAPAEKAAAAGAIETMAYELGAGLGIAIFGLLLSRSFSASIRLPA 423

Qy        270 TLRRLFVAELCSWMALMTFTLFYTDFVGEG--LYQGVP--RAEPGTEARRH---YDEGVR 322
              |    :|  | |            ||  |  :|  : :   :| ||   :   |
Db        424 GLEAQEIARASSSM-----------GEAVQLANSLPPTQGQAILDAARHAFIWSHSVA 470

Qy        323 MGSLG-LFLQCAISLVFSL 340
              : | | : |  |: : |||
Db        471 LSSAGSMLLLLAVGMWFSL 489


RESULT 15
Y051_MYCTU
ID   Y051_MYCTU     STANDARD;      PRT;   640 AA.
AC   Q10880; O53628;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein Rv0083/MT0090.
GN   RV0083 OR MT0090 OR MTCY251.01 OR MTV030.27.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
```

```
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RX   MEDLINE=22206494; PubMed=12218036;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT   "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   J. Bacteriol. 184:5479-5490(2002).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC       OF CHLOROPLASTS OR MITOCHONDRIA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021428; CAA16264.1; -.
DR   EMBL; AE006920; AAK44315.1; -.
DR   PIR; D70850; D70850.
DR   TIGR; MT0090; -.
DR   TubercuList; Rv0083; -.
DR   InterPro; IPR003918; NADHub_oxred4.
DR   InterPro; IPR001750; Oxidored_q1.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   PRINTS; PR01437; NUOXDRDTASE4.
KW   Hypothetical protein; Oxidoreductase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     52     72       POTENTIAL.
FT   TRANSMEM     90    110       POTENTIAL.
FT   TRANSMEM    136    156       POTENTIAL.
FT   TRANSMEM    179    199       POTENTIAL.
FT   TRANSMEM    208    228       POTENTIAL.
FT   TRANSMEM    241    261       POTENTIAL.
FT   TRANSMEM    277    297       POTENTIAL.
FT   TRANSMEM    298    318       POTENTIAL.
FT   TRANSMEM    352    372       POTENTIAL.
FT   TRANSMEM    391    411       POTENTIAL.
FT   TRANSMEM    446    466       POTENTIAL.
FT   TRANSMEM    497    517       POTENTIAL.
FT   TRANSMEM    619    639       POTENTIAL.
SQ   SEQUENCE   640 AA;  65627 MW;  6254C289DBD108A8 CRC64;

  Query Match             3.9%;  Score 111.5;  DB 1;  Length 640;
  Best Local Similarity   22.3%;  Pred. No. 0.89;
  Matches  129;  Conservative   73;  Mismatches  222;  Indels  155;  Gaps   29;

Qy         16 AQLLLV----NLLTFGL--EVCLAAGITYV-----PPLLLEVGVEEKFMTMVLGIGPVLG 64
              | :|||    :: || |  |:   | :  |       | :   |:    || :  |  ::|
Db         93 AAMLLVPAAGSVTTFLLAWELMAIASLILVLSEHARPQVRSAGLWYAVMTQLGFIAILVG 152

Qy         65 LVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLF-----------LIPRAGWLAGL 112
              || :   |  || : |        : |:  |:  ::|           |:|   ||
Db        153 LVVLAAAG-GSDRPAG--------LGAVCDGVRAAVFMLTLVGFGSKAGLVPLHAWLPRA 203

Qy        113 LCPDPRP----LELALLILGV-GLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFM 167
                | |    :  |:: ||: |:: |  |:           |   |     |      :
Db        204 HPEAPSPVSALMSAAMVNLGIYGIVRFDLQL-----------LGPGPRWWGLA------L 246

Qy        168 ISLGG---CLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALG 224
              :::||     | |  ::  |   |  |  |:     ||:||     : |  | |: | ||
Db        247 LAVGGTSALYGVLQASVAADLKRLLAYSTTEN---MGLITL----ALGAATLFADTGAYG 299

Qy        225 PTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPR--------LHQLCCRMPRTLRRLFV 276
              |   |   :   :  |        |||    |::|          |  |  |||  |
Db        300 PASIAAAAAMLHMIAHA--AFKSLAFMAAGSVLAATGLRDLDLLGGLARRMPAT------ 351

Qy        277 AELCSWMALMTFTLFY-------------TDFVGEG-LYQGVPRAEPGTEARRHYDEGVR 322
                          |:|:               || |  | |  :  | ||  :       :
Db        352 ------------TVFFGVAALGACGLPLGAGFVSEWLLVQSLIHAAPGHDPIVALTTPLA 399

Qy        323 MGSLGLFLQCAISLVFS------LVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVA 376
              :| : |   ::: :         |  |  |   :  |  :|||| ||  :||
Db        400 VGVVALATGLSVAAMTKAFGIGFLARPRSTQAEAAREAPASMRAGMAIAAGA-CLVLAVA 458

Qy        377 --VVTASAALTGFTFSALQILPYT-LASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLP 433
                :|         |  | | : :| | ::      | ||  |             |  |
Db        459 PLLVAPMVRRAAATLPAAQAVKFTGLGAV------VRLPAMSG-------------SIAP 499

Qy        434 GPKPGAPFPNGHVGAGGSGLLP-------PPPA-----LCGASACDVSVRVVVGEPTEAR 481
              |     |       |    :|       | ||      |||:  |::||:     : |
Db        500 GVIAAAVL----AAALAVAVLARWRFRRRPAPARLPLWACGAA--DLTVRMQYTATSFAE 553

Qy        482 VVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSV 520
                | : :  |:   |:   ::  | | :|   :    :|
Db        554 --PLQRVFGDVLRPDTDIEVTHTAESRYMAERITYRTAV 590


Search completed: August 25, 2004, 02:27:43
Job time : 33.5196 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 25, 2004, 02:22:38 ; Search time 149.566 Seconds
(without alignments)
1166.587 Million cell updates/sec

Title: US-09-593-793A-113
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2861	100.0	553	4	Q96JT2	Q96jt2 homo sapien
2	2798	97.8	553	6	Q95KI5	Q95ki5 macaca fasc
3	2602	90.9	553	11	Q8K0H7	Q8k0h7 mus musculu
4	2531	88.5	501	6	Q95KC5	Q95kc5 macaca fasc
5	2099	73.4	450	11	Q8K252	Q8k252 mus musculu
6	519.5	18.2	599	5	Q9VSV1	Q9vsv1 drosophila
7	512	17.9	112	11	Q8R1I0	Q8r1i0 mus musculu
8	484.5	16.9	530	11	Q8C204	Q8c204 mus musculu
9	456	15.9	576	13	Q90Z74	Q90z74 oryzias lat
10	446	15.6	751	11	Q8K4S3	Q8k4s3 rattus norv
11	394.5	13.8	754	5	Q869P1	Q869p1 dictyosteli
12	347.5	12.1	515	10	O65803	O65803 daucus caro
13	347.5	12.1	515	10	Q9FNR6	Q9fnr6 daucus caro
14	346.5	12.1	501	10	Q9SQK6	Q9sqk6 vitis vinif
15	346.5	12.1	515	10	Q9SLN7	Q9sln7 daucus caro
16	346	12.1	612	10	Q9SQK5	Q9sqk5 vitis vinif
17	344.5	12.0	520	10	Q7XA53	Q7xa53 glycine max
18	338	11.8	512	10	Q9S7Z5	Q9s7z5 apium grave
19	337.5	11.8	523	10	O04077	O04077 vicia faba
20	336.5	11.8	501	10	Q9SP63	Q9sp63 vitis vinif
21	333.5	11.7	599	10	Q7X9X8	Q7x9x8 plantago ma
22	333	11.6	512	10	Q9ZTB9	Q9ztb9 apium grave
23	332.5	11.6	508	10	Q8VYX3	Q8vyx3 brassica ol
24	331	11.6	511	10	Q84RQ3	Q84rq3 lotus japon
25	330	11.5	516	10	Q43653	Q43653 solanum tub
26	329.5	11.5	524	10	Q9XHL6	Q9xhl6 pisum sativ
27	328.5	11.5	506	10	Q944W2	Q944w2 oryza sativ
28	328.5	11.5	594	10	O80605	O80605 arabidopsis
29	327.5	11.4	506	10	Q948L0	Q948l0 oryza sativ
30	325.5	11.4	604	10	Q9FVL6	Q9fvl6 lycopersico
31	324.5	11.3	502	10	Q9SP14	Q9sp14 alonsoa mer
32	324.5	11.3	512	10	Q8RWQ6	Q8rwq6 arabidopsis
33	323.5	11.3	512	10	Q39231	Q39231 arabidopsis
34	323.5	11.3	512	10	O80550	O80550 arabidopsis
35	323.5	11.3	595	10	Q84KR4	Q84kr4 oryza sativ
36	317.5	11.1	507	10	Q40583	Q40583 nicotiana t
37	317.5	11.1	530	10	Q9M535	Q9m535 euphorbia e
38	316.5	11.1	500	10	Q9FVJ6	Q9fvj6 lycopersico
39	313.5	11.0	605	10	Q7Y1Q0	Q7y1q0 solanum tub
40	313	10.9	521	10	Q9SXM0	Q9sxm0 zea mays (m
41	311	10.9	533	10	Q41152	Q41152 ricinus com
42	309.5	10.8	607	10	Q8LPM4	Q8lpm4 citrus sine
43	307.5	10.7	488	10	Q9FV92	Q9fv92 solanum tub
44	306.5	10.7	513	10	Q8VYX4	Q8vyx4 brassica ol
45	306.5	10.7	528	10	Q8LPM6	Q8lpm6 citrus sine

ALIGNMENTS

RESULT 1
Q96JT2
ID Q96JT2 PRELIMINARY; PRT; 553 AA.
AC Q96JT2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Prostein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=21139094; PubMed=11245466;
RA Xu J., Kalos M., Stolk J.A., Zasloff E.J., Zhang X., Houghton R.L.,
RA Filho A.M., Nolasco M., Badaro R., Reed S.G.;
RT "Identification and characterization of prostein, a novel prostate-
RT specific protein.";
RL Cancer Res. 61:1563-1568(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral Nervous System;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

```
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral Nervous System;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY033593; AAK54386.1; -.
DR   EMBL; BC050416; AAH50416.1; -.
SQ   SEQUENCE   553 AA;  59322 MW;  0AFA23FBC742A667 CRC64;

  Query Match           100.0%;  Score 2861;  DB 4;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 3.4e-210;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYSA 553
              |||||||||||||
Db        541 VVFDKSDLAKYSA 553


RESULT 2
Q95KI5
ID   Q95KI5      PRELIMINARY;      PRT;   553 AA.
AC   Q95KI5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Temporal cortex;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB060851; BAB46871.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   553 AA;  59392 MW;  0718F3A91FB3BF1E CRC64;

  Query Match           97.8%;  Score 2798;  DB 6;  Length 553;
  Best Local Similarity  98.2%;  Pred. No. 2.2e-205;
  Matches  542;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db          1 MVQRLWVSRLLRHRKAQLLLINLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PVLGLVSVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||  |
Db        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPAEPAEGLSAPSLPSH 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCPCWARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 YQGVPRAELGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHRERQVFLPKYRGDAGG 420
              |||||||||||||||||||||||||||||||||||||||||:|||||||||| ||
Db        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHRERQVFLPKYRGDAGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TSSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYS 552
              ||||||||||||
Db        541 VVFDKSDLAKYS 552


RESULT 3
Q8K0H7
ID   Q8K0H7      PRELIMINARY;      PRT;   553 AA.
AC   Q8K0H7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (PROSTEIN homolog).
GN   PROSTEIN OR 2210413P12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
```

```
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Urinary bladder;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; BC031381; AAH31381.1; -.
DR   EMBL; AK035428; BAC29063.1; -.
DR   MGD; MGI:1922082; Prostein.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
DR   GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR005805; Rieske.
DR   PROSITE; PS00200; RIESKE_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   553 AA;  59742 MW;  6A474FEA630B4B13 CRC64;

  Query Match           90.9%;  Score 2602;  DB 11;  Length 553;
  Best Local Similarity   90.9%;  Pred. No. 2e-190;
  Matches  502;  Conservative   14;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
              |:|||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MIQRLWASRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy         61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
              |||||| |||||||| |||||||||||||||||:||||||||||||||| || |||
Db         61 PVLGLVSVPLLGSASDQWRGRYGRRRPFIWALSLGVLLSLFLIPRAGWLAGLLYPDTRPL 120

Qy        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
              ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db        121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAFSVYAFMISLGGCLGYLLPA 180

Qy        181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
              |||||| |||||||||||||||||||| |:|||| | ||| ||| |||||    ::|
Db        181 IDWDTSVLAPYLGTQEECLFGLLTLIFLICMAATLFVTEEAVLGPPEPAEGLLVSAVSRR 240

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
              ||||   ||||||| | ||| |||||||||||||||||||||||||||||||||||||
Db        241 CCPCHVGLAFRNLGTLFPRLQQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
              ||||||||||||||||||:|||||||||||||||||||||:||||:||||||
Db        301 YQGVPRAEPGTEARRHYDEGIRMGSLGLFLQCAISLVFSLVMDRLVQKFGTRSVYLASVM 360

Qy        361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
              ||||| |||||||| |||||||||||||||||||||||||||||||||||||||||||| ||
Db        361 TFPVAAAATCLSHSVVVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDAGG 420

Qy        421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
              :| |||  |||||||||| |||||||:| ||:| |||||||||||||:||||||| ||
Db        421 SSGEDSQTTSFLPGPKPGALFPNGHVGSGSSGILAPPPALCGASACDVSMRVVVGEPPEA 480

Qy        481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
              ||| |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        481 RVVTGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSHSVTAYMVSAAGLGLVAIYFATQ 540

Qy        541 VVFDKSDLAKYS 552
              |||||:||||||
Db        541 VVFDKNDLAKYS 552


RESULT 4
Q95KC5
ID   Q95KC5      PRELIMINARY;      PRT;   501 AA.
AC   Q95KC5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Medulla oblongata;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB062977; BAB60745.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   501 AA;  53447 MW;  8C554BBD04EE0470 CRC64;

  Query Match           88.5%;  Score 2531;  DB 6;  Length 501;
  Best Local Similarity   97.8%;  Pred. No. 4.8e-185;
  Matches  489;  Conservative    2;  Mismatches    9;  Indels    0;  Gaps    0;

Qy         53 MTMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL 112
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db          1 MTMVLGIGPVLGLVSVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL 60

Qy        113 LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGG 172
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db         61 LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYTFMISLGG 120

Qy        173 CLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGL 232
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        121 CLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPAEPAEGL 180

Qy        233 SAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFY 292
              |||||  ||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SAPSLPSHCCPCWARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFY 240

Qy        293 TDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTR 352
              |||||||||||||| |||||||||||||||||||||||||||||:||||||||||
Db        241 TDFVGEGLYQGVPRAELGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVVDRLVQRFGTR 300

Qy        353 AVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLP 412
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db        301 AVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHRERQVFLP 360

Qy        413 KYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRV 472
              ||||| || |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KYRGDAGGTSSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRV 420

Qy        473 VVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGL 532
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 VVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGL 480

Qy        533 VAIYFATQVVFDKSDLAKYS 552
              ||||||||||||||||||||
Db        481 VAIYFATQVVFDKSDLAKYS 500


RESULT 5
Q8K252
ID   Q8K252      PRELIMINARY;      PRT;   450 AA.
AC   Q8K252;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   PROSTEIN OR 2210413P12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC034084; AAH34084.1; -.
DR   MGD; MGI:1922082; Prostein.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0045285; C:ubiquinol-cytochrome-c reductase complex; IEA.
DR   GO; GO:0008121; F:ubiquinol-cytochrome-c reductase activity; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   InterPro; IPR005805; Rieske.
DR   PROSITE; PS00200; RIESKE_2; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   450 AA;  48261 MW;  72C703AEC131302C CRC64;

  Query Match           73.4%;  Score 2099;  DB 11;  Length 450;
  Best Local Similarity   90.0%;  Pred. No. 4.1e-152;
  Matches  404;  Conservative   12;  Mismatches   33;  Indels    0;  Gaps    0;

Qy        104 PRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSV 163
              |||||||||| || |||||||||||||||||||||||||||||||||||||||||||:||
Db          1 PRAGWLAGLLYPDTRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAFSV 60

Qy        164 YAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAAL 223
              |||||||||||||||||||| |||||||||||||||||||| |:|||| | ||| |
Db         61 YAFMISLGGCLGYLLPAIDWDTSVLAPYLGTQEECLFGLLTLIFLICMAATLFVTEEAVL 120

Qy        224 GPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWM 283
              || ||||||   ::|  ||||   ||||||| | ||| ||||||||||||||||||||||
Db        121 GPPEPAEGLLVSAVSRRCCPCHVGLAFRNLGTLFPRLQQLCCRMPRTLRRLFVAELCSWM 180

Qy        284 ALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMD 343
              |||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        181 ALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGIRMGSLGLFLQCAISLVFSLVMD 240

Qy        344 RLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLY 403
              ||||:||||:|||||  ||||| |||||||| ||||||||||||||||||||||||||||
Db        241 RLVQKFGTRSVYLASVMTFPVAAAATCLSHSVVVVTASAALTGFTFSALQILPYTLASLY 300

Qy        404 HREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGA 463
              |||||||||||||| ||:| |||  |||||||||| |||||||:| ||:| ||||||||
Db        301 HREKQVFLPKYRGDAGGSSGEDSQTTSFLPGPKPGALFPNGHVGSGSSGILAPPPALCGA 360

Qy        464 SACDVSVRVVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAY 523
              ||||||:||||||| ||||| ||||||||||||||||||||||||||||||| |||||
Db        361 SACDVSMRVVVVGEPPEARVVTGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSHSVTAY 420

Qy        524 MVSAAGLGLVAIYFATQVVFDKSDLAKYS 552
              |||||||||||||||||||||:||||||
Db        421 MVSAAGLGLVAIYFATQVVFDKNDLAKYS 449


RESULT 6
Q9VSV1
ID   Q9VSV1      PRELIMINARY;      PRT;   599 AA.
AC   Q9VSV1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG4484 protein.
GN   CG4484.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003552; AAF50310.1; -.
DR   FlyBase; FBgn0035968; CG4484.
SQ   SEQUENCE   599 AA;  66057 MW;  C5381D334CFBF2EB CRC64;

  Query Match           18.2%;  Score 519.5;  DB 5;  Length 599;
  Best Local Similarity   23.7%;  Pred. No. 2.3e-31;
  Matches  150;  Conservative  100;  Mismatches  201;  Indels  183;  Gaps   14;

Qy         12 RHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLL 71
              | : ::  :: :   :|   ||  ::| |:||::||: | |:|  |: |::|    |||
Db         46 RKTRFEMFRLSAIAMAIEFAYAAETSFVSPILLQIGVDHKHMSMTWGLSPLIGFFMSPLL 105

Qy         72 GSASDHWRGRYGRRRPFIWALSLGILLSLFLIP-------------------------R 105
              || ||  : |:|||||  |  || ||:  | |:|
Db        106 GSISDRCKLRWGRRRPIISILSFGIMCGLILVPYGKDLGLLLGDAGYTYAESALNFTSSS 165

Qy        106 AGWLAGLLC------PDPRPLELA--LLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHC 157
              | :| |:       |     :|  | ||: ||||    | ||    | |:   |:
Db        166 GGSVAALVSGEATTGPSASDYKFAVILTILGMVLLDFDADTCQTPARTYLLDMC-VPEEQ 224

Qy        158 RQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATL-- 215
              :| :::|     || :|| :  :||:|: :  ::|     :| |:|:||  |   |:
Db        225 PKAMTMFALFAGFGGTIGYAIGGVDWETTHIGSFMGGNIPTVFTLVTIIFAVCYLITVTT 284

Qy        216 -------LVAEEAALGP-----------------------------------TEPAEG 231
                    |: ::  | |                                   | :|
Db        285 FREIPLPLIEQDELLRPLSEQAIKKELKKKNNTIYYIQETTQLELQMASDDPKRLEALQG 344

Qy        232 LSAPSLSP---------------HCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLF 275
                  ||              : |  :  :::                || ::| |
Db        345 SYQNGYSPAVEKQGKSQDLETQSDYDAPVSLKAYLKSIFI-----------MPYSMRMLA 393

Qy        276 VAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAIS 335
              :  |  ||  :|: |::|||||| :: | | | | :||  :|: ||| |  |: :
Db        394 LTNLFCWMGHVTYCLYFTDFVGEAVFHGDPTAAPNSEAALNYEAGVRFGCWGMAIYAFSC 453
```

```
Qy        336 LVFSLVMDRLVQRFGTRAVYLASVAAFPV---AAGATCLSHSVAVVTASAALTGFTFSAL 392
Db        454 SIYSLSVTKLMKWFGTKAVYISGMIYYGIGMLVLGLWPTKWGVLVFSTSA---GILYGTI 510

Qy        393 QILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSG 452
Db        511 FTVPFILVARYH-AKNCFSIK----------------------------NG-------- 532

Qy        453 LLPPPPALCGASACDVSVRVVVGEPTEARVVP---GRGICLDLAILDSAFLLSQVAPSLF 509
Db        533 --------------------------EIVPLKQARGLGTDVAIISSMVFIAQLIVSLS 564

Qy        510 MGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVF 543
Db        565 VGPLVSWMDTTCAVLYASTFLSFLAAIAAMFVLY 598


RESULT 7
Q8R1I0
ID   Q8R1I0      PRELIMINARY;      PRT;   112 AA.
AC   Q8R1I0;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC024519; AAH24519.1; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   112 AA;  11393 MW;  B9D9DE79FEF09FB7 CRC64;

  Query Match             17.9%;  Score 512;  DB 11;  Length 112;
  Best Local Similarity   91.9%;  Pred. No. 1.3e-31;
  Matches  102;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        442 PNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLL 501
Db          1 PNGHVGSGSSGILAPPPALCGASACDVSMRVVVGEPPEARVVTGRGICLDLAILDSAFLL 60

Qy        502 SQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVFDKSDLAKYS 552
Db         61 SQVAPSLFMGSIVQLSHSVTAYMVSAAGLGLVAIYFATQVVFDKNDLAKYS 111


RESULT 8
Q8C204
ID   Q8C204      PRELIMINARY;      PRT;   530 AA.
AC   Q8C204;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Membrane associated transporter protein.
GN   MATP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK089932; BAC41003.1; -.
DR   MGD; MGI:2153040; Matp.
SQ   SEQUENCE   530 AA;  57935 MW;  426898973B7D6420 CRC64;

  Query Match             16.9%;  Score 484.5;  DB 11;  Length 530;
  Best Local Similarity   26.1%;  Pred. No. 9.3e-29;
  Matches  146;  Conservative   84;  Mismatches  229;  Indels  101;  Gaps   12;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
Db         34 RLVMHSMAMFGREFCYAVEAAYVTPGLLSVGLPKSLYSMVWLLSPILGFLLQPVVGSASD 93

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRP---LELALLILGVGLLD 133
Db         94 HCRARWGRRRPYILTLAIMMLLGMALYLNGDAVVSALVANPRQKLIWAISITMVGVVLFD 153

Qy        134 FCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLG 193
Db        154 FSADFIDGPIKAYLFDVCSHQDK-EKGLHYHALFTGFGGALGYILGAIDWVHLDLGRLLG 212

Qy        194 TQEECLFGLLTLIFLTCVAATLLVAEEAAL--GPTEP-------AEGLSAPSLSPHCCPC 244
Db        213 TEFQVMFFFSALVLILCFITHLCSIPEAPLRDAATDPPSQQDPQGSSLSASGMHEY---- 268

Qy        245 RARLAFRNLGA----------------------LLPRLHQLCCRMPRTLRRLFVAELCSW 282
Db        269 GSIEKVKNGGADTEQPVQEWKNQKPSGQSQRTMSMKSLLRALVNMPSHYRCLCVSHLIGW 328

Qy        283 MALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVM 342
Db        329 TAFLSNMLFFTDFMGQIVYHGDPYGAHNSTEFLIYERGVEVGCWGLCINSVFSSVYSYFQ 388

Qy        343 DRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASL 402
Db        389 KAMVSYIGLKGLYFMGYLLFGLGTGFIGLFPNVYSTLVLCSMFGVMSSTLYTVPFNLIAE 448

Qy        403 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCG 462
Db        449 YHREEE----KEKGQEA----------------PGGPDNQGR----GQGV------DCA 477

Qy        463 ASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTA 522
Db        478 ALTCMVQL---------AQILVGGG----------------------LGFLVNMAGSVVV 506

Qy        523 YMVSAAGLGLVAIYFATQVV 542
Db        507 VVITASAVSLIGCCFVALFV 526


RESULT 9
Q90Z74
ID   Q90Z74      PRELIMINARY;      PRT;   576 AA.
AC   Q90Z74;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Membrane-associated transporter protein B.
OS   Oryzias latipes (Medaka fish) (Japanese ricefish).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
OX   NCBI_TaxID=8090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HNI;
RX   MEDLINE=21372467; PubMed=11479596;
RA   Fukamachi S., Shimada A., Shima A.;
RT   "Mutations in the gene encoding B, a novel transporter protein, reduce
RT   melanin content in medaka.";
RL   Nat. Genet. 28:381-385(2001).
DR   EMBL; AF332510; AAK77024.1; -.
SQ   SEQUENCE   576 AA;  63411 MW;  13A5C9828ABEE894 CRC64;
```

```
  Query Match           15.9%;  Score 456;  DB 13;  Length 576;
  Best Local Similarity   24.0%;  Pred. No. 1.5e-26;
  Matches  144;  Conservative   95;  Mismatches  198;  Indels  164;  Gaps   17;

Qy         12 RHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLL 71
              |  : :|:| ::: || | | |    :| |:|| ||:     ::|  | |:|| :  |::
Db         56 RRSRGRLILHSMVMFGREFCYAVEAAFVTPVLLSVGLPRSLYSLVWLISPILGFLLQPII 115

Qy         72 GSASDHWRGRYGRRRPFIWALSLGIL----LSLFL------------IPRAGWLAGLLC 114
              |||||: |  :|||||:|  | ||||    ||:||                |: |
Db        116 GSASDYCRSSWGRRRPYI--LVLGILMLVGLSMFLNGDAVVSELVSDRSSRSTW------ 167

Qy        115 PDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDL--FRDPDHCRQAYSVYAFMISLGG 172
                : ::: || | ||       |::| | |:  ::| :   :    :|    |||
Db        168 ------AIVVVMFGVVLFDFAADFIDGPIKAYLFDVCSYQDKE---RGLHYHALFTGLGG 218

Qy        173 CLGYLLPAIDWDTSALAPYLGTQEECLF-------GLLTLIFLTCV-------------- 211
              |||: |:||  | |    ||:: : ::       |:  :: |  :
Db        219 ACGYLVGAMDWGHSVLGRLLGSEYQVIYFFSALTWGVFLIVHLFSIPEKPLAKVPSESSA 278

Qy        212 -AATLLVAEEA----ALG--PTEPAEGLSAPSLSPHCCPC-------------------- 244
              :|  |:   :    |||  |  |     |:| : |
Db        279 SSALRLLGPHSNGYGALGKEPVSPVIPTSSPEIRPRSYSALGERPRSFSALGEANSVTSS 338

Qy        245 --------RARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFV 296
                  : :: ||:        | :    ||   | | :: |  | | :   ||:|||:
Db        339 AKQPIKEDQKKMTFRS-------LMKAIFNMPNHYRFLCISHLLGWAAFLCNMLFFTDFM 391

Qy        297 GEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYL 356
              |: :|:| | ||  : |   |: || :|  || :    | ::| |   |:  | : :|
Db        392 GQIVYRGNPYAEHNSTAYITYERGVEVGCWGLCINAVSSALYSYVQRFLLPYIGLKGLYF 451

Qy        357 ASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRG 416
                | :      |   |      :: |   | |  :|: | : | ||::  : | |
Db        452 MGYFVFGMGTSLIGLFPEVIATLILCSVPGVMSSTLYTIPFNLIAEYQREEEEQV-KLEG 510

Qy        417 DTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGE 476
                                    |:    |:|:       | |  | | :
Db        511 --------------------------GNESPRGTGM------DCAALTCMVQL------ 531

Qy        477 PTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIY 536
                |::: | |                          :|::|  : ||   ::||: : |:
Db        532 ---AQIIVGAG---------------------LGALVNAAGSVIVVVLSASSISLIGCI 566

Qy        537 F 537
              |
Db        567 F 567
```

```
RESULT 10
Q8K4S3
ID   Q8K4S3      PRELIMINARY;      PRT;   751 AA.
AC   Q8K4S3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Proton-associated sugar transporter A.
GN   PAST-A.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Shimokawa N., Okada J., Haglund K., Dikic I., Koibuchi N., Miura M.;
RT   "Past-A, a new class proton-associated sugar transport facilitator in
RT   the central nervous system.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB075229; BAB97313.1; -.
DR   GO; GO:0005351; F:sugar porter activity; IEA.
KW   Sugar transport.
```

```
SQ   SEQUENCE   751 AA;  81751 MW;  5C33E1C2E4A1B2EC CRC64;

  Query Match           15.6%;  Score 446;  DB 11;  Length 751;
  Best Local Similarity   22.3%;  Pred. No. 1.2e-25;
  Matches  160;  Conservative   79;  Mismatches  206;  Indels  272;  Gaps   15;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
              :||    : ||:|    |    || |:||::|: ::  ::|  | |:|| :  ||||: ||
Db         86 ELLFNGCILFGIEFSYAMETAYVTPVLLQMGLPDQLYSLVWFISPILGFLLQPLLGAWSD 145

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDP---RPLELALLILGVGLLD 133
                 |:|||||||  |::| || | |:   |   |:   |          : | : || |:|
Db        146 RCTSRFGRRRPFILVLAIGALLGLSLLLN-GRDIGMALADTATNHKWGILLTVCGVVLMD 204

Qy        134 FCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLG 193
              |       |  | : |:    |  |   :::| |  |||  ||::  | || ::    ||
Db        205 FSADSADNPSHAYMMDVCGPVDQDR-GLNIHALMAGLGGGFGYVVGGIHWDKTSFGRALG 263

Qy        194 TQEECLFGLLTLIFLTCVAATLLVA-----------------------------EEAA 222
              |   :: : | | |:      ||:                              || |
Db        264 GQLRVIY-IFTAITLSVTTVFTLVSIPERPLRPLGEKRTAMKSPSLPLPPSPPVLLEEGA 322

Qy        223 LGPTEP---AEGLSAPSLSPHCCP---------------------------------- 243
              | | |   |  | |   ||   |
Db        323 -GDTLPSTTATSLYASFSSPISPPSPLTPKYGSFISRDSSLTGINEFASSFGTSNIDSVL 381

Qy        244 ----------------------------------CRARLAFRNLGALL----------- 257
                                               |: |  |  | | |
Db        382 IDCFTAGHDNYLALPSSVPRQAISVSFPRAPDGFYCQERGLERREGPLTLGLDGDVLRVG 441

Qy        258 ------PR---------------------------------------------------- 259
                    ||
Db        442 SLDTSKPRASGILKRPQTLALPDVAGGNGPETSRRRNVTFSQQVANILLNGVKYESELTG 501

Qy        260 ----------LHQLCC---RMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPR 306
                       | :||      ||| :| | |      |::     ||||||:|| ::|| |:
Db        502 SSEQSEQPLSLRRLCSTIYNMPRPVRNLCVNHFLGWLSFEGMLLFYTDFMGEVVFQGDPK 561

Qy        307 AEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAA 366
              |   :|| : |: || ||  |: :    :  :| ::::| :    | :|  :    | :
Db        562 APHASEAYQKYNSGVTMGCWGMCIYAFSAAFYSAILEKLEECLSVRTLYFIAYLLFGLGT 621

Qy        367 GATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDS 426
              |   || :: || :    |  || |  |||:|   |:: |     |: |     || |
Db        622 GLATLSRNLYVVLSLCTHYGILFSTLCTLPYSLLCDYYQSK-----KFAG-----SSADG 671

Qy        427 LMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEARVVPGR 486
                                                                      |
Db        672 TR-------------------------------------------------------R 674

Qy        487 GICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVF 543
              |: :|:::|   : |:|:  || :|           : || |     :|||: | |
Db        675 GMGVDISLLSCQYFLAQILVSLVLGP----------LTSAVGSANGVMYFASLVSF 720
```

```
RESULT 11
Q869P1
ID   Q869P1      PRELIMINARY;      PRT;   754 AA.
AC   Q869P1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to Arabidopsis thaliana (Mouse-ear cress). sucrose transporter
DE   protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=22092622; PubMed=12097910;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
```

```
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL   Nature 418:79-85(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Baumgart C.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC117072; AAO52628.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR005828; Sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
SQ   SEQUENCE   754 AA;  82774 MW;  614C473C832ACCE4 CRC64;

  Query Match             13.8%;  Score 394.5;  DB 5;  Length 754;
  Best Local Similarity   23.5%;  Pred. No. 1e-21;
  Matches  128;  Conservative   90;  Mismatches  200;  Indels  127;  Gaps   13;

Qy         13 HRKA----QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCV 68
Db        197 HKKTLPLFNLICLTICFLGVQFGWALQIAFSTPLFLELGVEQKWVSYIWLAGPISGLIVQ 256

Qy         69 PLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL--CPDPRPLELALLI 126
Db        257 PLVGVITDRSECRFGRRKPFILIGSVFISIGLVLISNAETFGSYFGDSEQKKSIAISIAI 316

Qy        127 LGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTS 186
Db        317 VGFWILDLSNNAVQAPCRALLVDI-AAPSQQSLGSSLFSLMLGTGNLLGYMMGSI--DLV 373

Qy        187 ALAPYLGTQEECLFGLLTLIFLTCVAATL-LVAEEAALGPTEPAEGLSAPSLSPHCCPCR 245
Db        374 RMVPFMKTDTRALFTLSIMVLLFCVVMTLGFVTEEQYIRVNED-QSVENP---------- 422

Qy        246 ARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVP 305
Db        423 -----------LKTMFKGIVKMPTYLQRLCAVQFFSWIGWFSFVLFITTWVGVNVFGGDP 471

Qy        306 RAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVA 365
Db        472 NAPEYSDSRILFQDGVRWGSLSLTISSGITIAVSLLIPFLVKFIDMKYIYIGGNL----- 526

Qy        366 AGATCLSHSVAVVTAS-------AALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 418
Db        527 --LQCIFFALFYFVESKIGSLLLIASTGIPWAIVMILPFSIVGM---------------- 568

Qy        419 GGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPT 478
Db        569 -GVEDNES-----------------------SG------------------------- 577

Qy        479 EARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFA 538
Db        578 -----------LNIGTLNIFVVVPQMVVSLGIGLILDLSKGNVVYSLLT---GSVASFFA 623

Qy        539 TQVVF 543
Db        624 TLFCF 628


RESULT 12
O65803
ID   O65803      PRELIMINARY;      PRT;   515 AA.
AC   O65803;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Sucrose/H+ symporter.
GN   SUT2.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Daucinae;
OC   Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Namtaise; TISSUE=Root;
RX   MEDLINE=99063785; PubMed=9847123;
RA   Shakya R., Sturm A.;
RT   "Characterization of source- and sink-specific sucrose/H+ symporters
RT   from carrot.";
RL   Plant Physiol. 118:1473-1480(1998).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; Y16768; CAA76369.1; -.
DR   PIR; T14340; T14340.
DR   GO; GO:0019866; C:inner membrane; IEA.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0008515; F:sucrose transporter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0015770; P:sucrose transport; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005989; Suc7H_symport.
DR   Pfam; PF00083; sugar_tr; 1.
DR   TIGRFAMs; TIGR01301; GPH_sucrose; 1.
KW   Transmembrane.
SQ   SEQUENCE   515 AA;  54424 MW;  897B90657C9E243C CRC64;

  Query Match             12.1%;  Score 347.5;  DB 10;  Length 515;
  Best Local Similarity   25.9%;  Pred. No. 2.6e-18;
  Matches  131;  Conservative   80;  Mismatches  215;  Indels   79;  Gaps   17;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
Db         34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
Db         94 HCQSSFGRRRPFIASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIGF 150

Qy        130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
Db        151 WILDVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLF 210

Qy        179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
Db        211 PFS--KTHACDLYCANLKSCFIISIALLIIITVVALSVVRENS--GPPDDADAAEEP--- 263

Qy        239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
Db        264 ----PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGR 313

Qy        299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
Db        314 EIYGGT--AGQG----KLYDQGVRAGALGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWG 366

Qy        359 VAAFPVAAGATCL-----------SHSV--------AVVTASA----ALTGFTFSALQIL 395
Db        367 FVNFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSI 426

Qy        396 PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGA 448
Db        427 PFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISA 486

Qy        449 GGSGLLP----PPPALCGASACDVS 469
Db        487 AISGVLAIVLLPKPSKDAASKLSLS 511


RESULT 13
Q9FNR6
ID   Q9FNR6      PRELIMINARY;      PRT;   515 AA.
AC   Q9FNR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
```

```
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Sucrose/proton symporter.
GN   SUT2.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Daucinae;
OC   Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nantaise;
RA   Sturm A.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nantaise;
RA   Shakya R.;
RL   Thesis (2000), Department of Department of Botany,
RL   University of Basel, Basel, Switzerland.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AJ303199; CAC19689.1; -.
DR   GO; GO:0019866; C:inner membrane; IEA.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0008515; F:sucrose transporter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0015770; P:sucrose transport; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005989; Suc/H_symport.
DR   Pfam; PF00083; sugar_tr; 1.
DR   TIGRFAMs; TIGR01301; GPH_sucrose; 1.
KW   Transmembrane.
SQ   SEQUENCE   515 AA;  54425 MW;  9580978C726523DC CRC64;

  Query Match             12.1%;  Score 347.5;  DB 10;  Length 515;
  Best Local Similarity   25.9%;  Pred. No. 2.6e-18;
  Matches  131;  Conservative   80;  Mismatches  215;  Indels   79;  Gaps   17;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
              :|:||  :  |::   |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db         34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
              | :  :|||||| : :  : :|: ||   |: | :       :    :|  : : ::|
Db         94 HCQSSFGRRRPFIASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIGF 150

Qy        130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
             :||   :  |  |||:||   |    | | : |:| :::|  |||             |
Db        151 WILDVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLF 210

Qy        179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
             |    | |   |    : |    : |: :  | |  :| | :  || : |:    |
Db        211 PFS--KTHACDLYCANLKSCFIISIALLIIITVVALSVVRENS--GPPDDADAAEEP--- 263

Qy        239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
                 |   ::     | ||  |  |    || :  | :    :|:|   | || ||::|
Db        264 ----PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGR 313

Qy        299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
              :| |   |  |    : ||:||| |:||| |   :  : |: :: ||:  |   : |
Db        314 EIYGGT--AGKG----KLYDQGVRAGALGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWG 366

Qy        359 VAAFPVAAGATCL-----------SHSV--------AVVTASA----ALTGFTFSALQIL 395
               | :| |                 ||          | | |  :: |   |    :
Db        367 FVNFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSI 426

Qy        396 PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGA 448
              |: |||:|          |    |     :: | | ||              |  | : |
Db        427 PFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISA 486

Qy        449 GGSGLLP----PPPALCGASACDVS 469
```

```
                ||:|     | |:   ||   :|
Db        487 AISGVLAIVLLPKPSKDAASKLSLS 511


RESULT 14
Q9SQK6
ID   Q9SQK6      PRELIMINARY;      PRT;   501 AA.
AC   Q9SQK6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Putative sucrose transporter.
GN   VVSUC11.
OS   Vitis vinifera (Grape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   Vitaceae; Vitis.
OX   NCBI_TaxID=29760;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Shiraz; TISSUE=Fruit;
RA   Davies C., Wolf T., Robinson S.P.;
RT   "Three putative sucrose transporters are differentially expressed in
RT   grapevine tissues.";
RL   Plant Sci. 147:93-100(1999).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF021808; AAF08329.1; -.
DR   GO; GO:0019866; C:inner membrane; IEA.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0008515; F:sucrose transporter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0015770; P:sucrose transport; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005989; Suc/H_symport.
DR   Pfam; PF00083; sugar_tr; 1.
DR   TIGRFAMs; TIGR01301; GPH_sucrose; 1.
KW   Transmembrane.
SQ   SEQUENCE   501 AA;  53938 MW;  4D0D4DE2EF2F4BA8 CRC64;

  Query Match             12.1%;  Score 346.5;  DB 10;  Length 501;
  Best Local Similarity   27.0%;  Pred. No. 3e-18;
  Matches  153;  Conservative   83;  Mismatches  197;  Indels  133;  Gaps   22;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
              :|| |  :  |::   |  :: : | : |:|:   : :::   ||: ||:  ||:|  ||
Db         29 RLLRVASVACGIQFGWALQLSLLTPYVQELGIPHAWSSIIWLCGPLSGLLVQPLVGHLSD 88

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL--CPDPRPLELALLILGVGLLDF 134
                |:|||||| | :  |:::: :|  :  : |||    | ||  :|  ::|  |||
Db         89 RCNSRFGRRRPFIVAGATSIVVAVLIIGFSADIGGLLGDGADRRPRAVATFVVGFWLLDV 148

Qy        135 CGQVCFTPLEALLSDLFRDPDH--CRQAYSVYAFMISLGGCLGYLLPA------IDWDTS 186
                |   | |||:||  : ||   | | : ::  |::|  ||:   :        | | ||
Db        149 ANNVTQGPCRALLADL-TEKDHRRTRVANAYFSLFIAVGNVLGFATGSYSGWFRIFWFTS 207

Qy        187 ALAPYLGTQEECL----FGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCC 242
                :       :|        || :||   :| |  ::  ||          ||: | | |
Db        208 TSS----CNADCANLKSAFLLDIIF---IAITTYISITAA-----QELPLSSSSRSTHIS 255

Qy        243 -----PCRARLAFRNLGALLPRLHQLCCRMPRTLRR--------LFVAELCSWMALMTFT 289
                    |: ||          | :|      |||          |||  | :|: |: |
Db        256 EEMAESTHAQEAF---------LWELF----GTLRYLSGSIWIILFVTAL-TWIGLLPFL 301

Qy        290 LFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 349
              || ||::|  :| | |          ::|: |||||:||| |   :  :  |::|::| :::
Db        302 LFDTDWMGREIYGGKP------NEGQNYNTGVRMGALGLMLNSVVLGITSVLMEKLCRKW 355

Qy        350 GTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQV 409
              |   |:                |  |   |  : ||    | : | :
Db        356 GAGFVW----------------------GLSNILMSLCFLLMLILS---AVVKHMD--- 386

Qy        410 FLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVS 469
```

```
                                  ||    |    |:| | |          ||   |   :
Db        387 --------------------FLGHDLP----PSGVVIA----------ALIVFSILGIP 411

Qy        470 VRVVVGEP---TEARVVP-GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL-----SQSV 520
              : :    |       |:   | | | : :|: | :: ||  ||  |   ||      | |:
Db        412 LAITYSVPYALISTRIESLGLGQGLSMGVLNLAIVIPQVIVSLGSGPWDQLFGGGNSPSL 471

Qy        521 TAYMVSAAGLGLVAIYFATQVVFDKS 546
                |:|   |||||    :   |||
Db        472 AVAAVAAFASGLVAILAIPRSSADKS 497


RESULT 15
Q9SLN7
ID   Q9SLN7      PRELIMINARY;      PRT;   515 AA.
AC   Q9SLN7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Sucrose transporter protein.
GN   CSUT.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC   campanulids; Apiales; Apiaceae; Apioideae; Scandiceae; Daucinae;
OC   Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Root;
RA   Wu N., Diao F., Zhang L., Huang M.;
RT   "Sucrose transporter protein.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB036758; BAA89458.1; -.
DR   GO; GO:0019866; C:inner membrane; IEA.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0008515; F:sucrose transporter activity; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0015770; P:sucrose transport; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR005828; Sub_transporter.
DR   InterPro; IPR005989; Suc7H_symport.
DR   Pfam; PF00083; sugar_tr; 1.
DR   TIGRFAMs; TIGR01301; GPH_sucrose; 1.
KW   Transmembrane.
SQ   SEQUENCE   515 AA;  54523 MW;  2B548D1D9DFF51AC CRC64;

  Query Match             12.1%;  Score 346.5;  DB 10;  Length 515;
  Best Local Similarity   27.2%;  Pred. No. 3.1e-18;
  Matches  137;  Conservative   76;  Mismatches  215;  Indels   75;  Gaps   19;

Qy         17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
              :|:||  :  |::   |  :: : | :  :|:  :|:   :   ||: |::  |::|  ||
Db         34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy         77 HWRGRYGRRRPFIWALSLGILLSLFLIPRA---GWLAG-LLCPDPRPLELALLILGVGLL 132
              | :  :|||||| : :  : :|: ||  |   |: ||  :    :|  : : ::|  :|
Db         94 HCQSSFGRRRPFIASGAGCVAISVILIGFAADIGYKAGDDMSKTLKPRAVTVFVIGFWIL 153

Qy        133 DFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LLPAI 181
              |    :  |  |||:||   |    | | : |:| :::|  |||           | |
Db        154 DVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLFPFS 213

Qy        182 DWDTSALAPYLGTQEEC-LFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
                | |   |    : | :  :  || :| ||  | |  |    | : |:    |
Db        214 --KTHACDLYCANLKSCFIISIALLIIITVVA--LSVVREKQWSP-DDADAADEP----- 263

Qy        241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
                |   ::     | ||  |  |    || :  | :    :|:|   | || ||::|  :
Db        264 --PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGREI 315

Qy        301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
                | |   |  |    : ||:||| ||||| |   :  : |: :: ||:  |   : |  |
Db        316 YGGT--AGKG----KLYDQGVRAGSLGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWGVV 368

Qy        361 AFPVAAGATCL-----------SHSV--------AVVTASA----ALTGFTFSALQILPY 397
                | :| |                ||        | | | |    :: |   |    :|:
Db        369 NFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSIPF 428

Qy        398 TLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGAGG 450
                |||:|              |     |      :: | | ||             |  | : |
Db        429 ALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISAAI 488

Qy        451 SGLLP----PPPALCGASACDVS 469
                ||:|     | |:   ||   :|
Db        489 SGVLAIVLLPKPSKDAASKLSLS 511


Search completed: August 25, 2004, 02:31:18
Job time : 154.566 secs
```